

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number  
**WO 02/31152 A2**

(51) International Patent Classification<sup>7</sup>: C12N 15/12,  
C07K 14/47, A01K 67/027, C07K 16/18, C12Q 1/68,  
A61K 38/17, G01N 33/68, 33/53

(21) International Application Number: PCT/US01/32090

(22) International Filing Date: 12 October 2001 (12.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/240,871 13 October 2000 (13.10.2000) US  
60/244,723 30 October 2000 (30.10.2000) US  
60/249,402 15 November 2000 (15.11.2000) US  
60/252,622 22 November 2000 (22.11.2000) US  
60/255,622 13 December 2000 (13.12.2000) US

(71) Applicant (for all designated States except US): INCYTE  
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo  
Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BAUGHN, Mariah,  
R. [US/US]; 14244 Santiago Road, San Leandro, CA  
94577 (US). DING, Li [CN/US]; 3353 Alma Street, #146,  
Palo Alto, CA 94306 (US). ELLIOTT, Vicki, S. [US/US];  
3770 Polton Place Way, San Jose, CA 95121 (US).  
GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1,  
Menlo Park, CA 94025 (US). GIETZEN, Kimberly, J.  
[US/US]; 691 Los Huecos Drive, San Jose, CA 95123  
(US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way,  
Fremont, CA 94555 (US). GURURAJAN, Rajagopal  
[IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US).  
HAFALIA, April, J., A. [US/US]; 2227 Calle de Pri-  
mavera, Santa Clara, CA 95045 (US). KEARNEY, Liam  
[IE/US]; 50 Woodside Avenue, San Francisco, CA 94127  
(US). KHAN, Farrah, A. [IN/US]; 3617 Central Road,  
#102, Glenview, IL 60025 (US). LAL, Preeti [IN/US];  
P.O. Box 5124, Santa Clara, CA 95056 (US). LEE, Ernes-  
tine, A. [US/US]; 624 Kains Street, Albany, CA 94706  
(US). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San  
Jose, CA 95123 (US). LU, Yan [CN/US]; 3885 Corrina

Way, Palo Alto, CA 95123 (US). NGUYEN, Daniel, B.  
[US/US]; 1403 Ridgewood Drive, San Jose, CA 95118  
(US). ARVIZU, Chandra [US/US]; 490 Sherwood Way,  
#1, Menlo Park, CA 94025 (US). RAMKUMAR, Jaya  
[IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US).  
TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose,  
CA 95118 (US). THANGAVELU, Kavitha [IN/US];  
1950 Montecito Avenue, #23, Mountain View, CA 94043  
(US). THORNTON, Michael [US/US]; 9 Medway  
Road, Woodside, CA 94062 (US). WALIA, Narinder, K.  
[US/US]; 890 Davis Street, #205, San Leandro, CA 94577  
(US). WARREN, Bridget, A. [US/US]; 10130 Parkwood  
Drive, #2, Cupertino, CA 95014 (US). XU, Yuming  
[US/US]; 1739 Walnut Drive, Mountain View, CA 94040  
(US). YAO, Monique, G. [US/US]; 1189 Woodgate Drive,  
Carmel, IN 46033 (US). YUE, Henry [US/US]; 826 Lois  
Avenue, Sunnyvale, CA 94087 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,  
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

**Published:**

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: INTRACELLULAR SIGNALING MOLECULES

(57) Abstract: The invention provides human intracellular signaling molecules (INTSIG) and polynucleotides which identify and encode INTSIG. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of INTSIG.



WO 02/31152 A2

## INTRACELLULAR SIGNALING MOLECULES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of intracellular signaling molecules and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative, autoimmune/inflammatory, neurological, gastrointestinal, reproductive, and developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of intracellular signaling molecules.

### BACKGROUND OF THE INVENTION

Cell-cell communication is essential for the growth, development, and survival of multicellular organisms. Cells communicate by sending and receiving molecular signals. An example of a molecular signal is a growth factor, which binds and activates a specific transmembrane receptor on the surface of a target cell. The activated receptor transduces the signal intracellularly, thus initiating a cascade of biochemical reactions that ultimately affect gene transcription and cell cycle progression in the target cell.

Intracellular signaling is the process by which cells respond to extracellular signals (hormones, neurotransmitters, growth and differentiation factors, etc.) through a cascade of biochemical reactions that begins with the binding of a signaling molecule to a cell membrane receptor and ends with the activation of an intracellular target molecule. Intermediate steps in the process involve the activation of various cytoplasmic proteins by phosphorylation via protein kinases, and their deactivation by protein phosphatases, and the eventual translocation of some of these activated proteins to the cell nucleus where the transcription of specific genes is triggered. The intracellular signaling process regulates all types of cell functions including cell proliferation, cell differentiation, and gene transcription, and involves a diversity of molecules including protein kinases and phosphatases, and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens that regulate protein phosphorylation.

Cells also respond to changing conditions by switching off signals. Many signal transduction proteins are short-lived and rapidly targeted for degradation by covalent ligation to ubiquitin, a highly conserved small protein. Cells also maintain mechanisms to monitor changes in the concentration of denatured or unfolded proteins in membrane-bound extracytoplasmic compartments, including a transmembrane receptor that monitors the concentration of available chaperone molecules in the endoplasmic reticulum and transmits a signal to the cytosol to activate the transcription of nuclear genes encoding chaperones in the endoplasmic reticulum.

Certain proteins in intracellular signaling pathways serve to link or cluster other proteins involved in the signaling cascade. These proteins are referred to as scaffold, anchoring, or adaptor proteins. (For review, see Pawson, T. and J.D. Scott (1997) *Science* 278:2075-2080.) As many intracellular signaling proteins such as protein kinases and phosphatases have relatively broad substrate specificities, the adaptors help to organize the component signaling proteins into specific biochemical pathways. Many of the above signaling molecules are characterized by the presence of particular domains that promote protein-protein interactions. A sampling of these domains is discussed below, along with other important intracellular messengers.

## 10 Intracellular Signaling Second Messenger Molecules

### Protein Phosphorylation

Protein kinases and phosphatases play a key role in the intracellular signaling process by controlling the phosphorylation and activation of various signaling proteins. The high energy phosphate for this reaction is generally transferred from the adenosine triphosphate molecule (ATP) to a particular protein by a protein kinase and removed from that protein by a protein phosphatase. Protein kinases are roughly divided into two groups: those that phosphorylate serine or threonine residues (serine/threonine kinases, STK) and those that phosphorylate tyrosine residues (protein tyrosine kinases, PTK). A few protein kinases have dual specificity for serine/threonine and tyrosine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family (Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Books, Vol I:7-20, Academic Press, San Diego, CA).

STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), involved in mediating hormone-induced cellular responses; calcium-calmodulin (CaM) dependent protein kinases, involved in regulation of smooth muscle contraction, glycogen breakdown, and neurotransmission; and the mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York, NY, pp. 416-431, 1887).

PTKs are divided into transmembrane, receptor PTKs and nontransmembrane, non-receptor PTKs. Transmembrane PTKs are receptors for most growth factors. Non-receptor PTKs lack transmembrane regions and, instead, form complexes with the intracellular regions of cell surface receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin) and antigen-specific receptors on T and B lymphocytes.

Many of these PTKs were first identified as the products of mutant oncogenes in cancer cells in which their activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs, and it is well known that cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493).

An additional family of protein kinases previously thought to exist only in prokaryotes is the histidine protein kinase family (HPK). HPKs bear little homology with mammalian STKs or PTKs but have distinctive sequence motifs of their own (Davie, J.R. et al. (1995) *J. Biol. Chem.* 270:19861-19867). A histidine residue in the N-terminal half of the molecule (region I) is an autophosphorylation site. Three additional motifs located in the C-terminal half of the molecule include an invariant asparagine residue in region II and two glycine-rich loops characteristic of nucleotide binding domains in regions III and IV. Recently a branched chain alpha-ketoacid dehydrogenase kinase has been found with characteristics of HPK in rat (Davie et al., *supra*).

Protein phosphatases regulate the effects of protein kinases by removing phosphate groups from molecules previously activated by kinases. The two principal categories of protein phosphatases are the protein (serine/threonine) phosphatases (PPs) and the protein tyrosine phosphatases (PTPs). PPs dephosphorylate phosphoserine/threonine residues and are important regulators of many cAMP-mediated hormone responses (Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508). PTPs reverse the effects of protein tyrosine kinases and play a significant role in cell cycle and cell signaling processes (Charbonneau and Tonks, *supra*). As previously noted, many PTKs are encoded by oncogenes, and oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This hypothesis is supported by studies showing that overexpression of PTPs can suppress transformation in cells, and that specific inhibition of PTPs can enhance cell transformation (Charbonneau and Tonks, *supra*).

#### Phospholipid and Inositol-phosphate Signaling

Inositol phospholipids (phosphoinositides) are involved in an intracellular signaling pathway that begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane to the biphosphate state (PIP<sub>2</sub>) by inositol kinases. Simultaneously, the G-protein linked receptor binding stimulates a trimeric G-protein which in turn activates a phosphoinositide-specific phospholipase C-β. Phospholipase C-β then cleaves PIP<sub>2</sub> into two products, inositol triphosphate (IP<sub>3</sub>) and diacylglycerol. These two products act as mediators for separate signaling events. IP<sub>3</sub> diffuses through the plasma membrane to induce calcium release from the endoplasmic reticulum (ER), while



diaacylglycerol remains in the membrane and helps activate protein kinase C, a serine-threonine kinase that phosphorylates selected proteins in the target cell. The calcium response initiated by IP<sub>3</sub> is terminated by the dephosphorylation of IP<sub>3</sub> by specific inositol phosphatases. Cellular responses that are mediated by this pathway are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

Oxysterols are oxygenated derivatives of cholesterol and have a wide range of biological activities. Oxysterols mediate cholesterol homeostasis, steroid biosynthesis and sphingolipid metabolism within the cell, but can also diffuse through the plasma membrane and act as extracellular messengers, affecting such processes as platelet aggregation, cell growth and apoptosis. Oxysterols interact with a number of receptors, including the oxysterol binding protein (OSBP), the sterol regulatory element binding protein, the cellular nucleic acid binding protein, the LXR nuclear hormone receptors, and the LDL receptor (for a review, see Schroepfer, G.J. (2000) *Physiol. Rev.* 80:361-554). OSBP is a high-affinity intracellular receptor for a variety of oxysterols that down-regulate cholesterol synthesis and stimulate cholesterol esterification. Upon ligand binding, OSBP translocates from the cytoplasm to the Golgi. This movement seems to be dependent on the presence of a pleckstrin homology domain (Lagace, T.A. et al. (1997) *Biochem. J.* 326:205-213). The oxysterol-induced apoptosis of leukemic T-cells seems to be mediated by OSBP occupancy (Bakos, J.T. et al. (1993) *J. Steroid Biochem. Mol. Biol.* 46:415-426).

#### Cyclic Nucleotide Signaling

Cyclic nucleotides (cAMP and cGMP) function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. In particular, cyclic-AMP dependent protein kinases (PKA) are thought to account for all of the effects of cAMP in most mammalian cells, including various hormone-induced cellular responses. Visual excitation and the phototransmission of light signals in the eye is controlled by cyclic-GMP regulated, Ca<sup>2+</sup>-specific channels. Because of the importance of cellular levels of cyclic nucleotides in mediating these various responses, regulating the synthesis and breakdown of cyclic nucleotides is an important matter. Thus adenylyl cyclase, which synthesizes cAMP from AMP, is activated to increase cAMP levels in muscle by binding of adrenaline to  $\beta$ -adrenergic receptors, while activation of guanylate cyclase and increased cGMP levels in photoreceptors leads to reopening of the Ca<sup>2+</sup>-specific channels and recovery of the dark state in the eye. In contrast, hydrolysis of cyclic nucleotides by cAMP and cGMP-specific phosphodiesterases (PDEs) produces the opposite of these and other effects mediated by increased cyclic nucleotide levels. PDEs appear to be particularly important in the regulation of cyclic nucleotides, considering the diversity found in this family of proteins. At least seven families of mammalian PDEs (PDE1-7) have been identified based on substrate specificity and affinity, sensitivity

to cofactors, and sensitivity to inhibitory drugs (Beavo, J.A. (1995) *Physiol. Rev.* 75:725-748). PDE inhibitors have been found to be particularly useful in treating various clinical disorders. Rolipram, a specific inhibitor of PDE4, has been used in the treatment of depression, and similar inhibitors are undergoing evaluation as anti-inflammatory agents. Theophylline is a nonspecific PDE inhibitor used in the treatment of bronchial asthma and other respiratory diseases (Banner, K.H. and C.P. Page (1995) *Eur. Respir. J.* 8:996-1000).

### G-Protein Signaling

Guanine nucleotide binding proteins (G-proteins) are critical mediators of signal transduction between a particular class of extracellular receptors, the G-protein coupled receptors (GPCRs), and intracellular second messengers such as cAMP and  $Ca^{2+}$ . G-proteins are linked to the cytosolic side of a GPCR such that activation of the GPCR by ligand binding stimulates binding of the G-protein to GTP, inducing an "active" state in the G-protein. In the active state, the G-protein acts as a signal to trigger other events in the cell such as the increase of cAMP levels or the release of  $Ca^{2+}$  into the cytosol from the ER, which, in turn, regulate phosphorylation and activation of other intracellular proteins. Recycling of the G-protein to the inactive state involves hydrolysis of the bound GTP to GDP by a GTPase activity in the G-protein. (See Alberts, B. et al. (1994) Molecular Biology of the Cell Garland Publishing, Inc. New York, NY, pp.734-759.) Two structurally distinct classes of G-proteins are recognized: heterotrimeric G-proteins, consisting of three different subunits, and monomeric, low molecular weight (LMW), G-proteins consisting of a single polypeptide chain.

The three polypeptide subunits of heterotrimeric G-proteins are the  $\alpha$ ,  $\beta$ , and  $\gamma$  subunits. The  $\alpha$  subunit binds and hydrolyzes GTP. The  $\beta$  and  $\gamma$  subunits form a tight complex that anchors the protein to the inner side of the plasma membrane. The  $\beta$  subunits, also known as G- $\beta$  proteins or  $\beta$  transducins, contain seven tandem repeats of the WD-repeat sequence motif, a motif found in many proteins with regulatory functions. Mutations and variant expression of  $\beta$  transducin proteins are linked with various disorders (Neer, E.J. et al. (1994) *Nature* 371:297-300; Margottin, F. et al. (1998) *Mol. Cell.* 1:565-574).

LMW GTP-proteins are GTPases which regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. They consist of single polypeptides which, like the  $\alpha$  subunit of the heterotrimeric G-proteins, are able to bind and hydrolyze GTP, thus cycling between an inactive and an active state. At least sixty members of the LMW G-protein superfamily have been identified and are currently grouped into the six subfamilies of ras, rho, arf, sar1, ran, and rab. Activated ras genes were initially found in human cancers, and subsequent studies confirmed that ras function is critical in determining whether cells continue to grow or become differentiated. Other members of the LMW G-protein superfamily have roles in signal transduction that vary with the

function of the activated genes and the locations of the G-proteins.

Guanine nucleotide exchange factors regulate the activities of LMW G-proteins by determining whether GTP or GDP is bound. GTPase-activating protein (GAP) binds to GTP-ras and induces it to hydrolyze GTP to GDP. In contrast, guanine nucleotide releasing protein (GNRP) binds to GDP-ras and induces the release of GDP and the binding of GTP.

Other regulators of G-protein signaling (RGS) also exist that act primarily by negatively regulating the G-protein pathway by an unknown mechanism (Druey, K.M. et al. (1996) *Nature* 379:742-746). Some 15 members of the RGS family have been identified. RGS family members are related structurally through similarities in an approximately 120 amino acid region termed the RGS domain and functionally by their ability to inhibit the interleukin (cytokine) induction of MAP kinase in cultured mammalian 293T cells (Druey et al., *supra*).

#### Calcium Signaling Molecules

$\text{Ca}^{2+}$  is another second messenger molecule that is even more widely used as an intracellular mediator than cAMP.  $\text{Ca}^{2+}$  can enter the cytosol by two pathways, in response to extracellular signals. One pathway acts primarily in nerve signal transduction where  $\text{Ca}^{2+}$  enters a nerve terminal through a voltage-gated  $\text{Ca}^{2+}$  channel. The second is a more ubiquitous pathway in which  $\text{Ca}^{2+}$  is released from the ER into the cytosol in response to binding of an extracellular signaling molecule to a receptor.  $\text{Ca}^{2+}$  directly activates regulatory enzymes, such as protein kinase C, which trigger signal transduction pathways.  $\text{Ca}^{2+}$  also binds to specific  $\text{Ca}^{2+}$ -binding proteins (CBPs) such as calmodulin (CaM) which then activate multiple target proteins in the cell including enzymes, membrane transport pumps, and ion channels. CaM interactions are involved in a multitude of cellular processes including, but not limited to, gene regulation, DNA synthesis, cell cycle progression, mitosis, cytokinesis, cytoskeletal organization, muscle contraction, signal transduction, ion homeostasis, exocytosis, and metabolic regulation (Celio, M.R. et al. (1996) Guidebook to Calcium-binding Proteins, Oxford University Press, Oxford, UK, pp. 15-20). Some  $\text{Ca}^{2+}$  binding proteins are characterized by the presence of one or more EF-hand  $\text{Ca}^{2+}$  binding motifs, which are comprised of 12 amino acids flanked by  $\alpha$ -helices (Celio, *supra*). The regulation of CBPs has implications for the control of a variety of disorders. Calcineurin, a CaM-regulated protein phosphatase, is a target for inhibition by the immunosuppressive agents cyclosporin and FK506. This indicates the importance of calcineurin and CaM in the immune response and immune disorders (Schwaninger M. et al. (1993) *J. Biol Chem.* 268:23111-23115). The level of CaM is increased several-fold in tumors and tumor-derived cell lines for various types of cancer (Rasmussen, C.D. and A.R. Means (1989) *Trends Neurosci.* 12:433-438).

The annexins are a family of calcium-binding proteins that associate with the cell membrane (Towle, C.A. and B.V. Treadwell (1992) *J. Biol. Chem.* 267:5416-5423). Annexins reversibly bind to

negatively charged phospholipids (phosphatidylcholine and phosphatidylserine) in a calcium dependent manner. Annexins participate in various processes pertaining to signal transduction at the plasma membrane, including membrane-cytoskeleton interactions, phospholipase inhibition, anticoagulation, and membrane fusion. Annexins contain four to eight repeated segments of about 60 residues. Each repeat folds into five alpha helices wound into a right-handed superhelix.

### Signaling Complex Protein Domains

PDZ domains were named for three proteins in which this domain was initially discovered. These proteins include PSD-95 (postsynaptic density 95), Dlg (*Drosophila* lethal(1)discs large-1), and ZO-1 (zonula occludens-1). These proteins play important roles in neuronal synaptic transmission, tumor suppression, and cell junction formation, respectively. Since the discovery of these proteins, over sixty additional PDZ-containing proteins have been identified in diverse prokaryotic and eukaryotic organisms. This domain has been implicated in receptor and ion channel clustering and in the targeting of multiprotein signaling complexes to specialized functional regions of the cytosolic face of the plasma membrane. (For a review of PDZ domain-containing proteins, see Ponting, C.P. et al. (1997) *Bioessays* 19:469-479.) A large proportion of PDZ domains are found in the eukaryotic MAGUK (membrane-associated guanylate kinase) protein family, members of which bind to the intracellular domains of receptors and channels. However, PDZ domains are also found in diverse membrane-localized proteins such as protein tyrosine phosphatases, serine/threonine kinases, G-protein cofactors, and synapse-associated proteins such as syntrophins and neuronal nitric oxide synthase (nNOS). Generally, about one to three PDZ domains are found in a given protein, although up to nine PDZ domains have been identified in a single protein. The glutamate receptor interacting protein (GRIP) contains seven PDZ domains. GRIP is an adaptor that links certain glutamate receptors to other proteins and may be responsible for the clustering of these receptors at excitatory synapses in the brain (Dong, H. et al. (1997) *Nature* 386:279-284). The *Drosophila* scribble (SCRIB) protein contains both multiple PDZ domains and leucine-rich repeats. SCRIB is located at the epithelial septate junction, which is analogous to the vertebrate tight junction, at the boundary of the apical and basolateral cell surface. SCRIB is involved in the distribution of apical proteins and correct placement of adherens junctions to the basolateral cell surface (Bilder, D. and N. Perrimon (2000) *Nature* 403:676-680).

The PX domain is an example of a domain specialized for promoting protein-protein interactions. The PX domain is found in sorting nexins and in a variety of other proteins, including the PhoX components of NADPH oxidase and the Cpk class of phosphatidylinositol 3-kinase. Most PX domains contain a polyproline motif which is characteristic of SH3 domain-binding proteins (Ponting, C.P. (1996) *Protein Sci.* 5:2353-2357). SH3 domain-mediated interactions involving the PhoX

components of NADPH oxidase play a role in the formation of the NADPH oxidase multi-protein complex (Leto, T.L. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:10650-10654; Wilson, L. et al. (1997) *Inflamm. Res.* 46:265-271).

The SH3 domain is defined by homology to a region of the proto-oncogene c-Src, a cytoplasmic protein tyrosine kinase. SH3 is a small domain of 50 to 60 amino acids that interacts with proline-rich ligands. SH3 domains are found in a variety of eukaryotic proteins involved in signal transduction, cell polarization, and membrane-cytoskeleton interactions. In some cases, SH3 domain-containing proteins interact directly with receptor tyrosine kinases. For example, the SLAP-130 protein is a substrate of the T-cell receptor (TCR) stimulated protein kinase. SLAP-130 interacts via its SH3 domain with the protein SLP-76 to affect the TCR-induced expression of interleukin-2 (Musci, M.A. et al. (1997) *J. Biol. Chem.* 272:11674-11677). Another recently identified SH3 domain protein is macrophage actin-associated tyrosine-phosphorylated protein (MAYP) which is phosphorylated during the response of macrophages to colony stimulating factor-1 (CSF-1) and is likely to play a role in regulating the CSF-1-induced reorganization of the actin cytoskeleton (Yeung, Y.-G. et al. (1998) *J. Biol. Chem.* 273:30638-30642). The structure of the SH3 domain is characterized by two antiparallel beta sheets packed against each other at right angles. This packing forms a hydrophobic pocket lined with residues that are highly conserved between different SH3 domains. This pocket makes critical hydrophobic contacts with proline residues in the ligand (Feng, S. et al. (1994) *Science* 266:1241-1247).

A novel domain, called the WW domain, resembles the SH3 domain in its ability to bind proline-rich ligands. This domain was originally discovered in dystrophin, a cytoskeletal protein with direct involvement in Duchenne muscular dystrophy (Bork, P. and M. Sudol (1994) *Trends Biochem. Sci.* 19:531-533). WW domains have since been discovered in a variety of intracellular signaling molecules involved in development, cell differentiation, and cell proliferation. The structure of the WW domain is composed of beta strands grouped around four conserved aromatic residues, generally tryptophan.

Like SH3, the SH2 domain is defined by homology to a region of c-Src. SH2 domains interact directly with phospho-tyrosine residues, thus providing an immediate mechanism for the regulation and transduction of receptor tyrosine kinase-mediated signaling pathways. For example, as many as ten distinct SH2 domains are capable of binding to phosphorylated tyrosine residues in the activated PDGF receptor, thereby providing a highly coordinated and finely tuned response to ligand-mediated receptor activation. (Reviewed in Schaffhausen, B. (1995) *Biochim. Biophys. Acta.* 1242:61-75.)

The GSG domain (GRP33, Sam68, GLD-1) and the KH domain (an RNA binding domain), are found within Sam68, a 68-kDa Src substrate associated during mitosis protein, which is an RNA-

binding protein with signaling properties. It is known to be a substrate for Src-family tyrosine kinases during mitosis and associates with various SH3 and SH2 domain-containing signaling molecules. SLM-1 and SLM-2 (Sam68-like mammalian) proteins have sequence identity with Sam68, also contain the GSG domain, have proline-rich motifs, arginine-glycine repeats, and a C-terminal tyrosine-rich region. SLM-1 is a Src substrate during mitosis, suggesting a possible involvement in the steps of mitosis. It has been suggested by Di Fruscio et al. that Sam68/SLM defines a family in which the members have the potential to link tyrosine kinase signaling cascades with some aspects of RNA metabolism, possibly as multifunctional adapter proteins during mitosis (Di Fruscio, M. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96:2710-2715.)

The pleckstrin homology (PH) domain was originally identified in pleckstrin, the predominant substrate for protein kinase C in platelets. Since its discovery, this domain has been identified in over 90 proteins involved in intracellular signaling or cytoskeletal organization. Proteins containing the pleckstrin homology domain include a variety of kinases, phospholipase-C isoforms, guanine nucleotide release factors, and GTPase activating proteins. For example, members of the FGD1 family contain both Rho-guanine nucleotide exchange factor (GEF) and PH domains, as well as a FYVE zinc finger domain. FGD1 is the gene responsible for faciogenital dysplasia, an inherited skeletal dysplasia (Pasteris, N.G. and J.L. Gorski (1999) Genomics 60:57-66). Many PH domain proteins function in association with the plasma membrane, and this association appears to be mediated by the PH domain itself. PH domains share a common structure composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. Variable loops connecting the component beta strands generally occur within a positively charged environment and may function as ligand binding sites (Lemmon, M.A. et al. (1996) Cell 85:621-624). Ankyrin (ANK) repeats mediate protein-protein interactions associated with diverse intracellular signaling functions. For example, ANK repeats are found in proteins involved in cell proliferation such as kinases, kinase inhibitors, tumor suppressors, and cell cycle control proteins. (See, for example, Kalus, W. et al. (1997) FEBS Lett. 401:127-132; Ferrante, A.W. et al. (1995) Proc. Natl. Acad. Sci. USA 92:1911-1915.) These proteins generally contain multiple ANK repeats, each composed of about 33 amino acids. Myotrophin is an ANK repeat protein that plays a key role in the development of cardiac hypertrophy, a contributing factor to many heart diseases. Structural studies show that the myotrophin ANK repeats, like other ANK repeats, each form a helix-turn-helix core preceded by a protruding "tip." These tips are of variable sequence and may play a role in protein-protein interactions. The helix-turn-helix region of the ANK repeats stack on top of one another and are stabilized by hydrophobic interactions (Yang, Y. et al. (1998) Structure 6:619-626).

The tetratricopeptide repeat (TPR) is a 34 amino acid repeated motif found in organisms from bacteria to humans. TPRs are predicted to form amphipathic helices, and appear to mediate protein-

protein interactions. TPR domains are found in CDC16, CDC23, and CDC27, members of the the anaphase promoting complex which targets proteins for degradation at the onset of anaphase. Other processes involving TPR proteins include cell cycle control, transcription repression, stress response, and protein kinase inhibition (Lamb, J.R. et al. (1995) Trends Biochem. Sci. 20:257-259).

5           The armadillo/beta-catenin repeat is a 42 amino acid motif which forms a superhelix of alpha helices when tandemly repeated. The structure of the armadillo repeat region from beta-catenin revealed a shallow groove of positive charge on one face of the superhelix, which is a potential binding surface. The armadillo repeats of beta-catenin, plakoglobin, and p120<sup>cas</sup> bind the cytoplasmic domains of cadherins. Beta-catenin/cadherin complexes are targets of regulatory signals that govern cell  
10   adhesion and mobility (Huber, A.H. et al. (1997) Cell 90:871-882).

          Eight tandem repeats of about 40 residues (WD-40 repeats), each containing a central Trp-Asp motif, make up beta-transducin (G-beta), which is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins). In higher eukaryotes G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues.

15           The discovery of new intracellular signaling molecules, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, neurological, gastrointestinal, reproductive, and developmental disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of intracellular signaling  
20   molecules.

## SUMMARY OF THE INVENTION

          The invention features purified polypeptides, intracellular signaling molecules, referred to collectively as "INTSIG" and individually as "INTSIG-1," "INTSIG-2," "INTSIG-3," "INTSIG-4,"  
25   "INTSIG-5," "INTSIG-6," "INTSIG-7," "INTSIG-8," "INTSIG-9," "INTSIG-10," "INTSIG-11," "INTSIG-12," "INTSIG-13," "INTSIG-14," "INTSIG-15," "INTSIG-16," "INTSIG-17," "INTSIG-18," "INTSIG-19," and "INTSIG-20." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring  
30   amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence

of SEQ ID NO:1-20.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-20. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:21-40.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the



group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

- 5           The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide  
10           comprises at least 60 contiguous nucleotides.

- Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ  
15           ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence  
20           complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

- 25           The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, c) a  
30           polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional INTS1G, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional INTS1G, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a

pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional INTSIG, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to  
5 a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an  
10 immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the  
15 activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and d) an  
20 immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of  
25 the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, the method  
30 comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20

contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

## BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"INTSIG" refers to the amino acid sequences of substantially purified INTSIG obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of INTSIG. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of INTSIG either by directly interacting with INTSIG or by acting on components of the biological pathway in which INTSIG participates.

An "allelic variant" is an alternative form of the gene encoding INTSIG. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to

allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding INTSIG include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as INTSIG or a polypeptide with at least one functional characteristic of INTSIG. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding INTSIG, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding INTSIG.

The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent INTSIG. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of INTSIG is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of INTSIG. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of INTSIG either by directly interacting with INTSIG or by acting on components of the biological pathway in which INTSIG participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind INTSIG polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired.

5 Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to  
10 immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a  
15 specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The  
20 nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a  
25 cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-  
30 handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA;

peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic INTSIG, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding INTSIG or fragments of INTSIG may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the



protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
5	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
10	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
15	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
20	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

25       Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

30       A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

      The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is  
35   one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

      A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

40       "Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a

diseased and a normal sample.

“Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures,<sup>9</sup> thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of INTSIG or the polynucleotide encoding INTSIG which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:21-40 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:21-40, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:21-40 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:21-40 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:21-40 and the region of SEQ ID NO:21-40 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-20 is encoded by a fragment of SEQ ID NO:21-40. A fragment of SEQ ID NO:1-20 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-20. For example, a fragment of SEQ ID NO:1-20 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-20. The precise length of a fragment of SEQ ID NO:1-20 and the region of SEQ ID NO:1-20 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a  
5 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e  
10 sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and “diagonals saved”=4. The “weighted” residue  
15 weight table is selected as the default. Percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from  
20 several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including “blastn,” that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called “BLAST 2 Sequences” that is used for direct pairwise comparison of two nucleotide sequences. “BLAST 2  
25 Sequences” can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

30 *Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example,  
5 as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,  
over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at  
least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous  
nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported  
by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a  
10 length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode  
similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes  
in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid  
sequences that all encode substantially the same protein.

15 The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to  
the percentage of residue matches between at least two polypeptide sequences aligned using a  
standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment  
methods take into account conservative amino acid substitutions. Such conservative substitutions,  
explained in more detail above, generally preserve the charge and hydrophobicity at the site of  
20 substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default  
parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e  
sequence alignment program (described and referenced above). For pairwise alignments of  
polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap  
25 penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default  
residue weight table. As with polynucleotide alignments, the percent identity is reported by  
CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise  
comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version  
30 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for  
example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989)

Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour.

5 Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, 10 such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid 15 sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

20 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect 25 cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of INTSIG which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of INTSIG which is useful in any of the antibody production methods disclosed herein or known in the 30 art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of INTSIG. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of INTSIG.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an INTSIG may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of INTSIG.

"Probe" refers to nucleic acid sequences encoding INTSIG, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the

specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the



artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence.

5 Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated  
10 regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent,  
15 chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose  
20 instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing INTSIG, nucleic acids encoding INTSIG, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

25 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the  
30 epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with

which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

10 "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

20 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

30 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at

least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human intracellular signaling molecules (INTSIG), the polynucleotides encoding INTSIG, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, neurological, gastrointestinal, reproductive, and developmental disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are intracellular signaling molecules. For example, SEQ ID NO:2 is 37% identical to *Schizosaccharomyces pombe* beta transducin (GenBank ID g3451308) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.1e-146, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a G-beta repeat WD40 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from MOTIFS analysis provides further corroborative evidence that SEQ ID NO:2 is a transducin.

In an alternative example, SEQ ID NO:6 is 85% identical to murine nedd-1 protein (GenBank ID g286103) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains a WD domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:6 is a protein involved in signal transduction.

In an alternative example, SEQ ID NO:10 is 51% identical to the human rho GTPase activating protein p115 (GenBank ID g840786) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $5.2e-211$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 contains a rhoGAP domain, an SH3 domain, and a Fes/CIP4 actin cytoskeleton regulatory protein domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The presence of these domains is confirmed by BLIMPS and MOTIFS analyses, providing further corroborative evidence that SEQ ID NO:10 is a GTPase activating protein.

In an alternative example, SEQ ID NO:16 is 49% identical to the human ras-related tumor suppressor NOEY2 (GenBank ID g4100355) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $3.6e-45$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also contains a ras family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:16 is a signaling protein of the ras family.

In an alternative example, SEQ ID NO:20 is 95% identical to murine SLM-1 protein (GenBank ID g4426613) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is  $3.1e-183$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:20 also contains a KH domain (E-value is 0.11) as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) SEQ ID NO:1, SEQ ID NO:3-5, SEQ ID NO:7-9, SEQ ID NO:11-13, SEQ ID NO:14-15, and SEQ ID NO:17-19 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-20 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:21-40 or that distinguish between SEQ ID NO:21-40 and

related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 105283R6 is the identification number of an Incyte cDNA sequence, and BMARNOT02 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71206562V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g3034305) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from

genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses INTSIG variants. A preferred INTSIG variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the INTSIG amino acid sequence, and which contains at least one functional or structural characteristic of INTSIG.

The invention also encompasses polynucleotides which encode INTSIG. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40, which encodes INTSIG. The polynucleotide sequences of SEQ ID NO:21-40, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding INTSIG. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least

about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding INTSIG. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:21-40 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:21-40. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of INTSIG.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding INTSIG, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring INTSIG, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode INTSIG and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring INTSIG under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding INTSIG or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding INTSIG and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode INTSIG and INTSIG derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding INTSIG or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:21-40 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-



511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding INTSIG may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length,

to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include  
5 sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary  
10 sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer  
15 controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode INTSIG may be cloned in recombinant DNA molecules that direct expression of INTSIG, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy  
20 of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express INTSIG.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter INTSIG-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA  
25 shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such  
30 as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of INTSIG, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is

produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding INTSIG may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7*:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7*:225-232.) Alternatively, INTSIG itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) *Proteins, Structures and Molecular Properties*, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of INTSIG, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active INTSIG, the nucleotide sequences encoding INTSIG or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding INTSIG. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding INTSIG. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding INTSIG and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional

or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding INTSIG and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding INTSIG. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding INTSIG. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding INTSIG can be achieved using a

multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding INTSIG into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for

5 *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of INTSIG are needed, e.g. for the production of antibodies, vectors which direct high level expression of INTSIG may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

10 Yeast expression systems may be used for production of INTSIG. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*;

15 Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of INTSIG. Transcription of sequences encoding INTSIG may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.

20 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., *The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill,

25 New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding INTSIG may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain

30 infective virus which expresses INTSIG in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of

DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

5 For long term production of recombinant proteins in mammalian systems, stable expression of INTSIG in cell lines is preferred. For example, sequences encoding INTSIG can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before  
10 being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These  
15 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat*  
20 confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins  
25 (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest  
30 is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding INTSIG is inserted within a marker gene sequence, transformed cells containing sequences encoding INTSIG can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding INTSIG under the control of a single promoter. Expression of the marker gene in response to induction or selection

usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding INTSIG and that express INTSIG may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR

5 amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of INTSIG using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and  
10 fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on INTSIG is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and  
15 Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding INTSIG include  
20 oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding INTSIG, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety  
25 of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding INTSIG may be cultured under  
30 conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode INTSIG may be designed to contain signal sequences which direct secretion of INTSIG through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding INTSIG may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric INTSIG protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of INTSIG activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the INTSIG encoding sequence and the heterologous protein sequence, so that INTSIG may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled INTSIG may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

INTSIG of the present invention or fragments thereof may be used to screen for compounds that specifically bind to INTSIG. At least one and up to a plurality of test compounds may be screened for specific binding to INTSIG. Examples of test compounds include antibodies,



oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of INTSIG, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which INTSIG binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express INTSIG, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing INTSIG or cell membrane fractions which contain INTSIG are then contacted with a test compound and binding, stimulation, or inhibition of activity of either INTSIG or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with INTSIG, either in solution or affixed to a solid support, and detecting the binding of INTSIG to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

INTSIG of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of INTSIG. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for INTSIG activity, wherein INTSIG is combined with at least one test compound, and the activity of INTSIG in the presence of a test compound is compared with the activity of INTSIG in the absence of the test compound. A change in the activity of INTSIG in the presence of the test compound is indicative of a compound that modulates the activity of INTSIG. Alternatively, a test compound is combined with an in vitro or cell-free system comprising INTSIG under conditions suitable for INTSIG activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of INTSIG may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding INTSIG or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example,

mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding INTSIG may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding INTSIG can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding INTSIG is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress INTSIG, e.g., by secreting INTSIG in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## **THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of INTSIG and intracellular signaling molecules. In addition, the expression of INTSIG is closely associated with brain and neurological tissues including thoracic dorsal root ganglion tissue, dermal tissue, reproductive tissue, digestive and hemic/immune tissue, diseased prostate tissue, and tumorous tissues including bladder, tongue, and testicular. Therefore, INTSIG appears to play a role in cell proliferative, autoimmune/inflammatory, neurological, gastrointestinal, reproductive, and developmental disorders. In the treatment of disorders associated with increased INTSIG expression or activity, it is desirable to decrease the expression or activity of INTSIG. In the treatment of disorders associated with decreased INTSIG expression or activity, it is desirable to increase the

expression or activity of INTSIG.

Therefore, in one embodiment, INTSIG or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of INTSIG. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous

system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),

5 akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis,

10 gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation,

15 gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of

20 pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic pregnancy, teratogenesis; cancer of the breast,

25 fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas,

30 cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary

keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

5 In another embodiment, a vector capable of expressing INTSIG or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of INTSIG including, but not limited to, those described above.

10 In a further embodiment, a composition comprising a substantially purified INTSIG in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of INTSIG including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of INTSIG may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of INTSIG including, but not limited to, those listed above.

15 In a further embodiment, an antagonist of INTSIG may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of INTSIG. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, neurological, gastrointestinal, reproductive, and developmental disorders described above. In one aspect, an antibody which specifically binds INTSIG may be used directly as an antagonist or  
20 indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express INTSIG.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding INTSIG may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of INTSIG including, but not limited to, those described above.

25 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the  
30 various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of INTSIG may be produced using methods which are generally known in the art. In particular, purified INTSIG may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind INTSIG. Antibodies to INTSIG may

also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

5 For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with INTSIG or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, 10 polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to INTSIG have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or 15 fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of INTSIG amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to INTSIG may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not 20 limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the 25 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce INTSIG-specific single 30 chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in

the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for INTSIG may also be generated. For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between INTSIG and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering INTSIG epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for INTSIG. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of INTSIG-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple INTSIG epitopes, represents the average affinity, or avidity, of the antibodies for INTSIG. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular INTSIG epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the INTSIG-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of INTSIG, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of INTSIG-antibody

complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding INTSIG, or any  
5 fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding INTSIG. Such technology is well known in the art, and antisense oligonucleotides or larger  
10 fragments can be designed from various locations along the coding or control regions of sequences encoding INTSIG. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence  
15 complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other  
20 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding INTSIG may be used for  
25 somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475),  
30 cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated



cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides  
 5 brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in INTSIG expression or regulation causes disease, the expression of INTSIG from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in  
 10 INTSIG are treated by constructing mammalian expression vectors encoding INTSIG and introducing these vectors by mechanical means into INTSIG-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem.  
 15 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of INTSIG include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA),  
 20 and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). INTSIG may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau  
 25 (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding INTSIG from a normal individual.

30 Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al.

(1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to INTSIG expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding INTSIG under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding INTSIG to cells which have one or more genetic abnormalities with respect to the expression of INTSIG. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding INTSIG to target cells which have one or more genetic abnormalities with

respect to the expression of INTSIG. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing INTSIG to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding INTSIG to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for INTSIG into the alphavirus genome in place of the capsid-coding region results in the production of a large number of INTSIG-coding RNAs and the synthesis of high levels of INTSIG in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of INTSIG into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA

transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding INTSIG.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding INTSIG. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages

within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

5           An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding INTSIG. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular  
10 chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased INTSIG expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding INTSIG may be therapeutically useful, and in the treatment of disorders associated with  
15 decreased INTSIG expression or activity, a compound which specifically promotes expression of the polynucleotide encoding INTSIG may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in  
20 altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding INTSIG is exposed to at least one test compound thus obtained. The sample  
25 may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding INTSIG are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding INTSIG. The amount of hybridization may be quantified, thus forming  
30 the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins,

D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified  
5 oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.

10 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and  
15 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's  
20 Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of INTSIG, antibodies to INTSIG, and mimetics, agonists, antagonists, or inhibitors of INTSIG.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical,  
25 sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and  
30 proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active

ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising INTSIG or fragments thereof. For example, liposome preparations  
5 containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, INTSIG or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

10 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

15 A therapeutically effective dose refers to that amount of active ingredient, for example INTSIG or fragments thereof, antibodies of INTSIG, and agonists, antagonists or inhibitors of INTSIG, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose  
20 lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity.  
25 The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the  
30 severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of

about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind INTSIG may be used for the diagnosis of disorders characterized by expression of INTSIG, or in assays to monitor patients being treated with INTSIG or agonists, antagonists, or inhibitors of INTSIG. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for INTSIG include methods which utilize the antibody and a label to detect INTSIG in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring INTSIG, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of INTSIG expression. Normal or standard values for INTSIG expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to INTSIG under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of INTSIG expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding INTSIG may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of INTSIG may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of INTSIG, and to monitor regulation of INTSIG levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding INTSIG or closely related molecules may be used to identify nucleic acid sequences which encode INTSIG. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding INTSIG, allelic variants, or related



sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the INTSIG encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:21-40 or from genomic sequences including promoters, enhancers, and introns of the INTSIG gene.

Means for producing specific hybridization probes for DNAs encoding INTSIG include the cloning of polynucleotide sequences encoding INTSIG or INTSIG derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding INTSIG may be used for the diagnosis of disorders associated with expression of INTSIG. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic

cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, ectopic

pregnancy, teratogenesis; cancer of the breast, fibrocystic breast disease, galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, gynecomastia, hypergonadotropic and hypogonadotropic hypogonadism, pseudohermaphroditism, azoospermia, premature ovarian failure, acrosin deficiency, delayed puberty, retrograde ejaculation and anejaculation, haemangioblastomas, cystsphaeochromocytomas, paraganglioma, cystadenomas of the epididymis, and endolymphatic sac tumours; and a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. The polynucleotide sequences encoding INTSIG may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered INTSIG expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding INTSIG may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding INTSIG may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding INTSIG in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of INTSIG, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding INTSIG, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified

polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,

- 5 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

- 10 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

- 15 Additional diagnostic uses for oligonucleotides designed from the sequences encoding INTSIG may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding INTSIG, or a fragment of a polynucleotide complementary to the polynucleotide encoding INTSIG, and will be employed under optimized conditions for identification of a specific gene or  
20 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

- In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding INTSIG may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease  
25 in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding INTSIG are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the  
30 secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual

overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the  
5 high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of INTSIG include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be  
10 accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray  
15 can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the  
20 activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

25 In another embodiment, INTSIG, fragments of INTSIG, or antibodies specific for INTSIG may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of  
30 gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of

transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

5 Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of  
10 pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a  
15 signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the  
20 rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released  
25 February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated  
30 biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are

5 analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl

10 sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or

15 untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the

20 present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for INTSIG to quantify the levels of INTSIG expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and

25 detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

30 Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which

alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated  
5 biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the  
10 present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared  
15 with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et  
20 al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding INTSIG may be  
25 used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific  
30 region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop



genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding INTSIG on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, INTSIG, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between INTSIG and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with INTSIG, or fragments thereof, and washed. Bound INTSIG is then detected by methods well known in the art. Purified INTSIG can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing

antibodies capable of binding INTSIG specifically compete with a test compound for binding INTSIG. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with INTSIG.

In additional embodiments, the nucleotide sequences which encode INTSIG may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, are expressly incorporated by reference herein: U.S. Ser. No. 60/240,871, U.S. Ser. No. 60/244,723, U.S. Ser. No. 60/249,402, U.S. Ser. No. 60/252,622, and U.S. Ser. No. 60/255,622.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the *LIFESEQ GOLD* database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra,

units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column

5 chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte  
10 Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo  
15 excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP  
96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1  
20 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using  
25 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such  
30 as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold

parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:21-40. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative intracellular signaling molecules were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode intracellular signaling molecules, the encoded polypeptides were analyzed by querying against PFAM models for intracellular signaling molecules. Potential intracellular signaling molecules were also identified by homology to Incyte cDNA sequences that had been annotated as intracellular signaling molecules. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

### "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

### "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

## **VI. Chromosomal Mapping of INTSIG Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:21-40 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:21-40 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:38 was mapped to chromosome 7 within the interval from 112.90 to 113.40 centiMorgans.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding INTSIG are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding INTSIG. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). In particular, SEQ ID NO:30 shows a strong association with neurological tissues. 1292 libraries present in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) isolated from 20 tissue types were examined. SEQ ID NO:30 was found in 73 libraries, 43 (59%) of which were isolated from



neurological tissues. Of 113 incidences of SEQ ID NO:30 in all libraries, 75 were in nervous system libraries. SEQ IN NO:30 is useful for distinguishing between nervous tissues and, for example, cardiovascular or endocrine tissues.

### VIII. Extension of INTSIG Encoding Polynucleotides

5 Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30  
10 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

15 High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters  
20 for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

25 The concentration of DNA in each well was determined by dispensing 100  $\mu$ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5  $\mu$ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the  
30 concentration of DNA. A 5  $\mu$ l to 10  $\mu$ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:21-40 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech).

An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16

hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

## 5 X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), *supra*).

10 Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) *Science*  
15 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The  
20 array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of  
25 complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and  
30 poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with

GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### 10 Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### 30 Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with

an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission

spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## **XI. Complementary Polynucleotides**

Sequences complementary to the INTSIG-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring INTSIG. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of INTSIG. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the INTSIG-encoding transcript.

## **XII. Expression of INTSIG**

Expression and purification of INTSIG is achieved using bacterial or virus-based expression systems. For expression of INTSIG in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express INTSIG upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of INTSIG in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding INTSIG by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, INTSIG is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from INTSIG at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified INTSIG obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

INTSIG function is assessed by expressing the sequences encoding INTSIG at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of INTSIG on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding INTSIG and either CD64 or CD64-GFP. CD64 and

CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art.

- 5 Expression of mRNA encoding INTSIG and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XIV. Production of INTSIG Specific Antibodies**

- INTSIG substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to  
10 immunize rabbits and to produce antibodies using standard protocols.

- Alternatively, the INTSIG amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well  
15 described in the art. (See, e.g., Ausubel, 1995, *supra*, ch. 11.)

- Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the  
20 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-INTSIG activity by, for example, binding the peptide or INTSIG to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring INTSIG Using Specific Antibodies**

- 25 Naturally occurring or recombinant INTSIG is substantially purified by immunoaffinity chromatography using antibodies specific for INTSIG. An immunoaffinity column is constructed by covalently coupling anti-INTSIG antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

- 30 Media containing INTSIG are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of INTSIG (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/INTSIG binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and INTSIG is collected.



**XVI. Identification of Molecules Which Interact with INTSIG**

INTSIG, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled INTSIG, washed, and any wells with labeled INTSIG complex are assayed. Data obtained using different concentrations of INTSIG are used to calculate values for the number, affinity, and association of INTSIG with the candidate molecules.

Alternatively, molecules interacting with INTSIG are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

INTSIG may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

**XVII. Demonstration of INTSIG Activity**

INTSIG activity is associated with its ability to form protein-protein complexes and is measured by its ability to regulate growth characteristics of NIH3T3 mouse fibroblast cells. A cDNA encoding INTSIG is subcloned into an appropriate eukaryotic expression vector. This vector is transfected into NIH3T3 cells using methods known in the art. Transfected cells are compared with non-transfected cells for the following quantifiable properties: growth in culture to high density, reduced attachment of cells to the substrate, altered cell morphology, and ability to induce tumors when injected into immunodeficient mice. The activity of INTSIG is proportional to the extent of increased growth or frequency of altered cell morphology in NIH3T3 cells transfected with INTSIG.

Alternatively, INTSIG activity is measured by binding of INTSIG to radiolabeled formin polypeptides containing the proline-rich region that specifically binds to SH3 containing proteins (Chan, D.C. et al. (1996) *EMBO J.* 15:1045-1054). Samples of INTSIG are run on SDS-PAGE gels, and transferred onto nitrocellulose by electroblotting. The blots are blocked for 1 hr at room temperature in TBST (137 mM NaCl, 2.7 mM KCl, 25 mM Tris (pH 8.0) and 0.1% Tween-20) containing non-fat dry milk. Blots are then incubated with TBST containing the radioactive formin polypeptide for 4 hrs to overnight. After washing the blots four times with TBST, the blots are exposed to autoradiographic film. Radioactivity is quantitated by cutting out the radioactive spots and counting them in a radioisotope counter. The amount of radioactivity recovered is proportional to the activity of INTSIG in the assay.

Alternatively, INTSIG protein kinase activity is measured by quantifying the phosphorylation

of an appropriate substrate in the presence of gamma-labeled  $^{32}\text{P}$ -ATP. INTSIG is incubated with the substrate,  $^{32}\text{P}$ -ATP, and an appropriate kinase buffer. The  $^{32}\text{P}$  incorporated into the product is separated from free  $^{32}\text{P}$ -ATP by electrophoresis, and the incorporated  $^{32}\text{P}$  is quantified using a beta radioisotope counter. The amount of incorporated  $^{32}\text{P}$  is proportional to the protein kinase activity of INTSIG in the assay. A determination of the specific amino acid residue phosphorylated by protein kinase activity is made by phosphoamino acid analysis of the hydrolyzed protein.

Alternatively, an assay for INTSIG protein phosphatase activity measures the hydrolysis of para-nitrophenyl phosphate (PNPP). INTSIG is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1%  $\beta$ -mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH, and the increase in light absorbance of the reaction mixture at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of INTSIG in the assay (Diamond, R.H. et al. (1994) Mol. Cell Biol. 14:3752-3762).

An alternative assay measures INTSIG-mediated G-protein signaling activity by monitoring the mobilization of  $\text{Ca}^{2+}$  as an indicator of the signal transduction pathway stimulation. (See, e.g., Grynkiewicz, G. et al. (1985) J. Biol. Chem. 260:3440; McColl, S. et al. (1993) J. Immunol. 150:4550-4555; and Aussel, C. et al. (1988) J. Immunol. 140:215-220). The assay requires preloading neutrophils or T cells with a fluorescent dye such as FURA-2 or BCECF (Universal Imaging Corp, Westchester PA) whose emission characteristics are altered by  $\text{Ca}^{++}$  binding. When the cells are exposed to one or more activating stimuli artificially (e.g., anti-CD3 antibody ligation of the T cell receptor) or physiologically (e.g., by allogeneic stimulation),  $\text{Ca}^{++}$  flux takes place. This flux can be observed and quantified by assaying the cells in a fluorometer or fluorescent activated cell sorter. Measurements of  $\text{Ca}^{++}$  flux are compared between cells in their normal state and those transfected with INTSIG. Increased  $\text{Ca}^{++}$  mobilization attributable to increased INTSIG concentration is proportional to INTSIG activity.

Alternatively, INTSIG activity is measured by binding of INTSIG to a substrate which recognizes WD-40 repeats, such as ElonginB, by coimmunoprecipitation (Kamura, T. et al. (1998) Genes Dev. 12:3872-3881). Briefly, epitope tagged substrate and INTSIG are mixed and immunoprecipitated with commercial antibody against the substrate tag. The reaction solution is run on SDS-PAGE and the presence of INTSIG visualized using an antibody to the INTSIG tag. Substrate binding is proportional to INTSIG activity.

Alternatively, INTSIG activity is measured by measuring oxysterol binding. Epitope-tagged INTSIG is incubated with a radio-labeled oxysterol ligand, such as  $^3\text{H}$ -25-hydroxycholesterol. INTSIG is collected by immunoprecipitation with a commercial antibody against the epitope, and bound

hydroxycholesterol quantitated by scintillation counting. INTSIG activity is proportional to the amount of ligand bound.

#### **XVIII. Assay to Detect INTSIG Binding to RNA**

The binding of INTSIG to RNA can be assayed using a solid phase RNA binding assay.

- 5 Hemagglutinin- (HA) tagged wild type and mutant INTSIG in pcDNA3 are transiently transfected into COS cells using LipofectAMINE reagent (Life Technologies, Inc.) for expression and analysis of RNA binding to multiple, simultaneously purified INTSIG proteins. Anti-HA immunoprecipitated INTSIG bound to protein G-Sepharose is incubated with 30 ng of <sup>32</sup>P-labeled G8-5 RNA in 30  $\mu$ l of RNA binding buffer containing 1  $\mu$ g/ $\mu$ l poly(C) at room temperature for 20 min. with occasional
- 10 shaking. The beads are then washed twice with 700  $\mu$ l of RNA binding buffer and resuspended in 20  $\mu$ l of SDS-polyacrylamide gel electrophoresis sample buffer. The protein and RNA were separated by 10% SDS-polyacrylamide gel electrophoresis. The RNA bands ran with a mobility equivalent to 25-35 kDa, and this part of the gel is cut out and dried for autoradiography. The upper part of the gel is transferred to a polyvinylidene difluoride membrane and blotted with anti-HA antibody to detect
- 15 HA-INTSIG (Lin, et al. (1997) J. Biol. Chem. 272:27274-27280).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments.

- 20 Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
105283	1	105283CD1	21	105283CB1
3350821	2	3350821CD1	22	3350821CB1
5876846	3	5876846CD1	23	5876846CB1
3560269	4	3560269CD1	24	3560269CB1
4596874	5	4596874CD1	25	4596874CB1
3594012	6	3594012CD1	26	3594012CB1
7482435	7	7482435CD1	27	7482435CB1
3882333	8	3882333CD1	28	3882333CB1
7482809	9	7482809CD1	29	7482809CB1
1739178	10	1739178CD1	30	1739178CB1
7473630	11	7473630CD1	31	7473630CB1
1431520	12	1431520CD1	32	1431520CB1
1916304	13	1916304CD1	33	1916304CB1
378504	14	378504CD1	34	378504CB1
5275371	15	5275371CD1	35	5275371CB1
490576	16	490576CD1	36	490576CB1
1417657	17	1417657CD1	37	1417657CB1
1773215	18	1773215CD1	38	1773215CB1
3036986	19	3036986CD1	39	3036986CB1
2041080	20	2041080CD1	40	2041080CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	105283CD1	g12082811	0	[fl][Gallus gallus] B cell phosphoinositide 3-kinase adaptor
2	3350821CD1	g3451308	2.10E-146	[Schizosaccharomyces pombe] beta transducin (Hargrave P.A. et al. (1993) Bioessays 15:43-50)
3	5876846CD1	g173479	1.00E-18	[Schizosaccharomyces pombe] sds22+ (protein phosphatase-1 regulatory protein) (Ohkura, H. and Yanagida, M. (1991) Cell 64:149-157)
4	3560269CD1	g7243701	1.10E-21	[Drosophila melanogaster] WDS (7-WD- repeat protein)
5	4596874CD1	g2407788	2.90E-58	[Dictyostelium discoideum] TipD (cell differentiation protein) (Stege, J.T. et al. (1999) Dev. Genet. 25:64-77)
6	3594012CD1	g286103	0	[Mus musculus] nedd-1 protein (Kumar, S. et al. (1992) Biochem. Biophys. Res. Commun. 185:1155-1161; Kumar, S. et al. (1994) J. Biol. Chem. 269:11318-11326)
7	7482435CD1	g4191594	1.00E-192	[Homo sapiens] protein serine/threonine phosphatase 4 regulatory subunit 1 (Kloeker, S. and Wadzinski, B.E. (1999) J. Biol. Chem. 274:5339-5347)
8	3882333CD1	g2145127	4.90E-10	[Mus musculus] p56lck-associated adaptor protein Lad
9	7482809CD1	g10953956	0	[fl][Homo sapiens] sorting nexin 16
10	1739178CD1	g14028714	0	[fl][Mus musculus] Rho GTPase- activating protein
11	7473630CD1	g14794726	0	[fl][Homo sapiens] CUB and sushi multiple domains 1 protein
12	1431520CD1	g2909372	9.40E-86	[Homo sapiens] small glutamine-rich tetraatricopeptide (SGT) (Kordes, E. et al. (1998) Genomics 52:90-94)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
13	1916304CD1	g4096360	3.30E-58	[Rattus norvegicus] CR16 (SH3 binding neuronal protein) (Weiler, M.C. et al. (1996) J. Mol. Neurosci. 7:203-215)
14	378504CD1	g5640145	4.60E-54	[Schizosaccharomyces pombe] oxysterol-binding protein family (Schroepfer Jr, G.J. (2000) Physiol. Rev. 80:361-554)
15	5275371CD1	g10086260	2.90E-21	[Zea mays] (AF250191) calmodulin-binding protein MPCBP (Safadi, F. et al. (2000) J. Biol. Chem. 275:35457-35470)
16	490576CD1	g4100355	3.60E-45	[Homo sapiens] NOEY2 (ras related tumor suppressor) (Yu, Y. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96:214-219)
17	1417657CD1	g2330828	6.30E-42	[Schizosaccharomyces pombe] hypothetical trp-asp repeats containing protein (similar to Homo sapiens, PEX7-HUMAN peroxisomal targeting signal 2 receptor) (Braverman, N. et al. (1997) Nat. Genet. 15:369-376)
18	1773215CD1	g9622151	7.50E-40	[Homo sapiens] TNF intracellular domain-interacting protein
20	2041080CD1	g4426613	3.10E-183	[Mus musculus] SLM-1 (Di Fruscio M. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96:2710-2715)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	105283CD1	805	S145 S149 S213 S219 S233 S245 S304 S33 S39 S54 S575 S592 S63 S640 S642 S667 S727 S731 S740 S785 T125 T151 T153 T208 T264 T377 T44 T497 T515 T610 T658 T671 Y195	N238 N243 N350 N728	Rgd cell attachment site R711-D713	MOTIFS
2	3350821CD1	957	S13 S191 S302 S44 S444 S461 S547 S61 S611 S651 S687 S7 S707 S859 S929 T157 T34 T415 T504 T563	N285 N421 N674 N699 N926 N934 N943	Amp_Binding L559-K570 G_Beta_Repeats L643-L657 Beta-transducin family Trp-Asp repeats signature g_beta_repeats.prf: D633-T680 WD domain, G-beta repeat WD40 K152-H188, E326-I367, G375-S414, R535-N573, R618-D656 G-protein beta WD-40 repeats PR00320C L643-L657 Beta G-protein (transducin) PR00319B I175-I189 TRPASP REPEATS PD023302: E58-I220 PD024370: H688-L868 PD023864: N332-Q471	MOTIFS MOTIFS PROFILES SCAN HMMER_PFAM BLIMPS_PRINTS BLIMPS_PRINTS BLAST_PRODUM
3	5876846CD1	274	S252 S256 S257 S36 S85 S92 T162 T31 Y100	N152 N34 N51 N73	Rgd cell attachment site R1074-D1076 Leucine Rich Repeat LR: K29-K50, N51-T72, N73-K94, K95-G116, S147-E168, N169-L190 Leucine-rich repeat signature PR00019B A71-I84	MOTIFS HMMER_PFAM BLIMPS_PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	3560269CD1	1144	S1127 S116 S170 S232 S245 S267 S268 S276 S283 S288 S292 S384 S390 S440 S454 S510 S546 S570 S643 S663 S672 S804 S93 S973 T1023 T1132 T137 T180 T22 T304 T308 T327 T352 T48 T533 T59 T60 T66 T67 T69 T702 T77 T78 T823 T94 T985 Y1090 Y218	N1000 N1065 N460 N686 N934	WD domain, G-beta repeat WD40: G644-K682, I685-K726, A734-N772, R835-N871 SH3 domain SH3: P1054-E1109 Src homology 3 (SH3) BL50002: A1058-D1076, K1095-S1108 Neutrophil cytosol factor (contains SH3 domains) PR00499: V1056-D1076, D1076-S1092, I1093-V1106 Beta G-protein (transducin) PR00319A L650-D666 SH3 domain signature PR00452 D1068-K1083, N1085-G1094, Q1097-E1109, P1054-A1064	HMIMER_PFAM HMIMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLIMPS_PRINTS BLIMPS_PRINTS
5	4596874CD1	513	S128 S262 T135 T146 T278 T339 T419 T473	N241 N267	G_Beta_Repeats L245-V259, I370-S384, A457-V471 WD domain, G-beta repeat WD40: T220-N258, L264-K302, A306-D344, P387-D424, V433-D470 Beta-transducin family Trp-Asp repeats signature g_beta_repeats.prf: A235-F282 G-protein beta WD-40 repeats PR00320B A331-L345 Beta G-protein (transducin) PR00319B A457-V471	MOTIFS HMIMER_PFAM PROFILES BLIMPS_PRINTS BLIMPS_PRINTS BLIMPS_BLOCKS HMIMER_PFAM
6	3594012CD1	667	S115 S121 S16 S17 S171 S187 S244 S245 S274 S322 S396 S403 S404 S412 S490 S503 S513 S532 S573 S575 S61 S638 S644 S70 T195 T259 T30 T408 T606	N105 N158 N169 N527 N553 N93	Trp-Asp (WD) repeat BL00678 T333-W343 WD domain, G-beta repeat: F35-V68, P77-D112, V118-S154, S161-D198, Y205-D242, L248-D283, P290-K324	BLIMPS_BLOCKS HMIMER_PFAM



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	(cont.)				G-protein beta WD-40 repeat signature PR00320:L99-L113	BLIMPS_PRINTS
					MEMBRANE; REPETITIVE	BLAST_DOMO
					DM00299 A53618 194-239:S201-K247	
					BETA-TRANSDUCIN FAMILY TRP-ASP REPEATS	BLAST_DOMO
					DM00005 A53618 108-150:S115-T157	
					BETA-TRANSDUCIN FAMILY TRP-ASP REPEATS	BLAST_DOMO
					DM00005 A53618 67-107:K74-K114	
					CORONIN	BLAST_DOMO
					DM00614 A53618 151-192:N158-N200	
					G_Beta_Repeats: L99-L113 L185-V199	MOTIFS
7	7482435CD1	897	S157 S322 S328 S338 S349 S357 S377 S454 S505 S517 S518 S576 S635 S723 S750 S763 S868 T100 T199 T25 T48 T553 T648 T85 Y691	N119 N538 N558	PHOSPHATASE SUBUNIT PP2A A PROTEIN REGULATORY REPEAT MULTIGENE FAMILY PD005088:C612-A878, L163-L282, N78-L297	BLAST_PRODUM
8	3882333CD1	454	S51 S124 S176 S202 S225 S235 S248 S261 S317 S375 T116 T164 T415 T353		Src homology domain 2: W347-H423	HMMER_PPFAM
					SH2 domain signature PR00401: W347-L361, P367-R377, Q412-E426	BLIMPS_PRINTS
9	7482809CD1	344	S222 S253 S259 S286 S299 S312 S321 S323 S330 S39 S56 S82 S87 T134 T146 T21 T232 T288	N38 N63	PX domain:D105-D214	HMMER_PPFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9 (cont)					PROTEIN PHOSPHOLIPASE 3 KINASE D SORTING NEXIN D2 CHROMOSOME PHOSPHOINOSITIDE P47PHOX PD003685:K123-L211 (e-value: 5.5e-07)	BLAST_PRODUM
10	1739178CD1	1115	S1045 S1047 S1060 S1107 S126 S230 S310 S327 S329 S35 S426 S516 S562 S712 S756 S775 S780 S828 S846 S847 S853 S884 S91 S952 S959 S96 S964 T1054 T1098 T1100 T122 T131 T296 T309 T333 T384 T402 T408 T482 T493 T619 T755 T8 T999 Y721 Y79	N1106 N209 N464 N521 N749	RhoGAP domain: P536-P688 GTPase activator protein PF00620B: D588-E604 PROTEIN GTPASE DOMAIN PD00930B: L639-L679 RHOGAP HEMATOPOIETIC P115 PROTEIN C1 GTPASE ACTIVATION SH3 PD042850: E149-I535 PROTEIN GTPASE DOMAIN SH2 ACTIVATION ZINC 3 KINASE SH3 PHOSPHATIDYLINOSITOL REGULATORY PD000780: I535-Q686, SH3 domain SH3: I763-Q817 Src homology 3 (SH3) domain BL50002A: A767-A785 Fes/CIP4 actin regulatory protein domain FCH: Q38-F136 PH DOMAIN DM00470 P98171 405-693: F418-I709 DM00470 P15882 109-331: A489-V698 DM00470 A43953 74-296: A489-V698 DM00470 Q03070 63-292: K518-I709	HMMER_PFAM BLIMPS_PFAM BLIMPS_PRODUM BLAST_PRODUM BLAST_PRODUM HMMER_PFAM BLIMPS_BLOCKS HMMER_PFAM BLAST_DOMO
11	7473630CD1	839	S104 S112 S146 S174 S24 S408 S451 S53 S560 S630 S72 S801 T100 T102 T122 T16 T20 T518 T599 T609 T694 T699 T720 T86 T9	N287 N342 N460 N532 N679 N769	Sushi domain (SCR repeat) sushi: C279-C335, C452-C509, C628-C685 Sushi domain proteins PF00084B: G471-Y482 CUB domain: C165-F271, C339-Y444, C513-Y617, C689-Y794	HMMER_PFAM BLIMPS_PFAM HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11 (cont.)					CLR/CLS REPEAT DM00162 P98069 418-529: A337-Y444, C689-Y794, C165-F271 DM00162 I49540 748-862: A337-Y444, C689-T795 DM00162 A57190 826-947: W328-Y444, C165-S273, W678-Y794 DM00162 P98063 755-862: L343-Y444, L693-T795 GLYCOPROTEIN DOMAIN EGF-LIKE PROTEIN SIGNAL PRECURSOR RECEPTOR INTRINSIC FACTOR B12 PD000165: C339-Y444, C689-Y794, C165-F271	BLAST_DOMO
12	1431520CD1	304	S188 S19 S298 S299 S3 S46 S77 T25	N186	TPR Domain TPR: A85-N118, A119-Y152, S153-N186 TPR REPEAT DM00408 S61991 98-247: K84-E202 DM00408 P31948 1-147: K84-E202 DM00408 P53041 24-181: K84-E202 DM00408 P15705 1-149: A85-N221 SGT SMALL GLUTAMINERICH TETRATRICOPEPTIDE PROTEIN PD012682: M1-S65 PD030464: L200-E302 PROTEIN REPEAT DOMAIN TPR PR00126A: G92-I112	HMIMER_PPFAM BLAST_DOMO BLAST_PRODOM BLTIMPS_PRODOM
13	1916304CD1	440	S117 S131 S140 S144 S155 S197 S272 S30 S308 S391 T388 T425 T54	N163 N270	Wiskott Aldrich syndrome scaffolding protein homology region 2 WH2: G36-V53 GO PROLINE; RICH; DM05534 S31719 1-122: M1-P116 H-A-P-P REPEAT DM08271 S25299 69-249: P192-P374, Y203-H382 DM08271 P13983 30-248: P157-L378 PROLINE RICH PROTEIN PD083803: S76-R216 PD052395: N263-D317 PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPIANCHOR BRAIN MAJOR PD001091: P177-R427	HMIMER_PPFAM BLAST_DOMO BLAST_PRODOM BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
14	378504CD1	747	S114 S116 S13 S184 S189 S207 S225 S248 S397 S407 S43 S435 S460 S48 S555 S637 S644 S674 S684 S688 T291 T420 T501 T523 T659 T699 T9 Y732	N31 N622 N680	Oxysterol-binding protein Oxysterol_B: G334-E747 Oxysterol-binding protein BL01013: G380-I415, V505-P514, R675-W718 Pleckstrin Homology domain PH: G63-Q155 OXYSTEROL-BINDING PROTEIN FAMILY DM01394 P38755   27-408: F504-P741, D351-S477 DM01394 Q02201   27-408: F504-E747, D347-S477 DM01394 P35843   1-390: D382-D728 DM01394 P35844   1-390: D382-D728 PROTEIN STEROL BIOSYNTHESIS INTERGENIC KES1 OXYSTEROLBINDING CHROMOSOME HES1 PD003744: S471-K719, D351-S488, Q632-K738 TPR Domain TPR: F326-D359, P399-D443, V691-C724, A478-N511, E623-S656, H657-G690, H725-S758 PROTEIN REPEAT DOMAIN TPR PD00126B: G664-L684	HMMER_PPFAM BLIMPS_BLOCKS HMMER_PPFAM BLAST_DOMO
15	5275371CD1	770	S206 S373 S389 S420 S462 S48 S559 S570 S586 S590 S595 T102 T110 T134 T417 T534 T565 T575 T620 T723 T750 Y17	N131 N132	Ras family ras: R9-M199 RAS TRANSFORMING PROTEIN DM00006 P10114   1-145: D7-T150 DM00006 P22123   1-145: D7-E149 DM00006 P10113   1-145: D7-E149 DM00006 A31961   1-145: D7-E149 GTP-binding nuclear protein ran proteins BL01115: Y8-D51, S90-S133, R141-R171 Transforming protein P21 ras signature PR00449: T31-V47, I48-A70, E112-E125, F147-E169, Y8-K29 ATP/GTP-binding site motif A (P-loop) Atp_Gtp-A: G14-S21 signal_cleavage: M1-R33	HMMER_PPFAM BLAST_DOMO BLIMPS_BLOCKS BLIMPS_PRINTS MOTIFS SPSCAN
16	490576CD1	199	S134 S151 S93 T142 T173 T31 T39 T43 T89			

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
17	1417657CD1	790	S72 S76 S155 S165 S172 S181 S364 S398 S503 S598 S600 S679 S727 T17 T239 T498 T522 T701 Y345	N74 N468 N691 N718	WD domain, G-beta repeat: R19-A53, V67-N103, Q112-D149, S155-D192, C199-D236, P289-L327 G-PROTEIN BETA WD-40 REPEAT PR00320: L223-M237 HYPOTHETICAL 93.2 KD TRPASP REPEATS CONTAINING PROTEIN C4F8.11 IN CHROMOSOME I REPEAT WD PD145764: T238-G784 Trp-Asp (WD) repeats signature: L90-L104, L223-M237, T269-V283 signal_cleavage: M1-E52 PH domain: V19-N119	HMME PFAM  BLAST_PRINTS  BLAST_PRODUM  MOTIFS  SPSCAN HMME PFAM
18	1773215CD1	490	S29 S217 S244 S256 S278 S318 S324 S356 S390 T103 T199 T250 T423			
19	3036986CD1	914	S66 S69 S314 S337 S344 S397 S406 S418 S543 S557 S568 S822 S834 S845 S870 T77 T265 T489 T715 T830 T886 Y503 Y601	N117 N494 N541 N864	TPR Domain: A668-F701, R702-H735, I736- N770, V771-E804, A446-D479, I480-I513, L528-F562, K563-N596, A597-H630 transmembrane domain: K238-Q258, F78-E202 F32D1.3 PROTEIN SIMILAR E NIDULANS BIMA GENE PRODUCT PD041324: M243-L415	HMME PFAM  HMME BLAST_PRODUM

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
20	2041080CD1	349	S54 S61 S87 S106 S184 S298 S328 S336 T222 T244 T324 Y49 Y124	N179	KH domain, R63-E115 (e=0.11) PHOSPHOPROTEIN P62 TYROSINE ASSOCIATED TSTAR ETOILE GAP ASSOCIATED SAM68 DELTA KH SRC MITOSIS PD016035: P216-Y349 PROTEIN PHOSPHOPROTEIN P62 ZF1 TYROSINE PUTATIVE TRANSCRIPTION FACTOR NUCLEAR GAP ASSOCIATED PD149659: I58-E115 PHOSPHOPROTEIN P62 TYROSINE ASSOCIATED TSTAR ETOILE GAP ASSOCIATED SAM68 DELTA KH SRC MITOSIS PD016104: E3-I58 PROTEIN ZF1 PUTATIVE PHOSPHOPROTEIN P62 TRANSCRIPTION FACTOR NUCLEAR KH RNA PD002056: G120-S181 do PHOSPHOPROTEIN; P62; GAP; RAS-GAP; DM02127 A38219 82-278: M1-G180 do PHOSPHOPROTEIN; P62; GAP; RAS-GAP; DM02127 I49140 82-278: M1-G180 do PHOSPHOPROTEIN; P62; GAP; RAS-GAP; DM02127 P13230 1-202: Y49-D162 do PHOSPHOPROTEIN; P62; GAP; RAS-GAP; DM02127 S52735 66-258: K59-S181	HMMER_PFAM BLAST_PRODUM BLAST_PRODUM BLAST_PRODUM BLAST_PRODUM BLAST_PRODUM BLAST_PRODUM

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
21	105283CB1	2860	1-370	105283R6 (EMARNOT02) 7278594H1 (EMARTXE01) 71206562V1 71205509V1 7723366H2 (THYRDIE01) 8268570H1 (LIVRTXF01) 7723366J2 (THYRDIE01) 72116818D1	916 1 2114 2112 1565 487 1235 1539	1386 536 2860 2601 2235 1145 1937 1984
22	3350821CB1	3542	1-87, 2036- 2055, 3291- 3542, 689- 1433	GNN.g6165165_018.edit 72080931D1 72082762D1 8004658H1 (PENIFEC01) 7179959H1 (BRAXDIC01) 72080558D1 GBI.g7341444_000001.edit5p 72072023V1 8120218H1 (TONSDIC01) 8036966H1 (SMCRUNE01)	460 2066 2652 832 1424 2021 1 2885 26 1314	1213 2831 3405 1439 2071 2753 710 3542 682 1929
23	5876846CB1	1014	1-65, 958- 1014	4211510F6 (BRONDIT01) 7937202H1 (CONNTMA01)	292 1	1014 684
24	3560269CB1	4040	1066-1891, 454-477, 3525-4040	6110780T8 (MCLDXTX03) 71013637V1 70075254U1 6123724T8 (BRAHNON05) 5957031H1 (BRATNOT05) 6123724F8 (BRAHNON05) 7230468H1 (BRAXTDR15) 2502132T6 (ADRETUT05) 663423R1 (BRAINT03) 6009953F6 (FIBRUNT02) 5588511F6 (ENDINOT02) 5510241F6 (BRADDIR01)	528 1 2191 1758 3331 1379 920 430 2741 3528 3079 2674	1260 519 2715 2369 4003 1927 1528 1057 3252 4040 3532 3206
25	4596874CB1	2006	1-453	71994085V1 72131979D1 72131721D1 72131460D1	753 640 1281 1	1475 1430 2006 747

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	3594012CB1	3643	2330-2365, 1- 912, 3177- 3643	1878258F6 (LEUKNOT03)	2378	2892
				70843402V1	1019	1666
				8081715U1	1584	2343
				8018226J1 (BMATXE01)	538	1208
				5718304F6 (PANCNOT16)	1	686
				1296328T6 (PGANNT03)	2621	3303
				2450413T6 (ENDANOT01)	3014	3643
				71222810V1	1910	2391
				3014319T6 (MUSCNOT07)	2237	2869
				71984885V1	2043	2694
27	7482435CB1	2694	405-534, 1- 60, 954-1071, 1650-1686, 1876-2259	71986615V1	1498	2203
				71985401V1	1323	2201
				72355610D1	608	1376
				72294113V1	724	1412
				4401241F6 (TESTTUT03)	1	648
				70929483V1	919	1454
				7080738H1 (STOMTMR02)	401	976
				71976518V1	1176	1986
				70931523V1	1467	2022
				71979346V1	1	821
28	3882333CB1	2349	1-26, 1711- 1834, 2282- 2349	71278610V1	1751	2349
				2079658F6 (UTRSNOT08)	775	1213
				2170258H1 (ENDCNOT03)	1	252
				2266454R6 (UTRSNOT02)	292	858
				7337260H1 (CONFEDN02)	39	662
				7101655H1 (BRAWTDR02)	2139	2674
				71303535V1	1139	1761
				3204864H1 (PENCNOT03)	2935	3201
				71156812V1	511	1136
				71156219V1	1119	1747
29	7482809CB1	1213	1137-1213	71303559V1	1710	2308
				71156491V1	1832	2325
				7284165H1 (BRAIFEJ01)	116	380
				5808761H2 (BRATNOT05)	3233	3465
30	1739178CB1	3465	2590-3465, 1591-1704, 711-1029, 1- 115, 2345- 2474			



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
30 (cont.)	1739178CB1	3465	2590-3465, 1591-1704, 711-1029, 1- 115, 2345- 2474	GNN.g5931375_002.edit 71156021V1 GNN.g7644424_000002_004.edi t	1 313 1513	471 892 3348
31	7473630CB1	2609	153-906, 1388-2609	GNN.g6648531_004.edit 58007679J1 56003273J1 GBI.g7342122_000016.edit 5877418F9 (BRAUNOT01) g3034305 GNN.g7243881_000003_002 GNN.g7243881_000004_002	90 1377 850 2476 2003 1 192 907	348 2284 1304 2609 2569 152 906 1620
32	1431520CB1	2580	1781-2580, 1- 23	5260422F6 (CONDTUT01) 7005922H1 (COLNFEC01) 7087982H1 (BRAUTDR03) 4063010F6 (BRAINOT21) 7447501F8 (BRAYDIN03) 7188091H1 (BRATDIC01) 5285811T9 (TESTNON04) 6314635H1 (NERDTDN03) 7462535H1 (LIVRFE04) 70158598V1 71042696V1 70175351V1 71040083V1 70174490V1	1602 1935 1424 992 1 2175 682 269 1 1308 437 1089 693 1606	2092 2514 1822 1743 659 2580 1242 996 548 1910 1038 1722 1211 2181
33	1916304CB1	2181	1-266, 2103- 2122, 1212- 1291	6202948H1 (PITUNON01) 6085421H1 (LUNLTUT11) 6875871H1 (EPIMUNN04) 426115R6 (BLADNOT01) 7080048H1 (STOMTMR02) 6608771H1 (BRSTTMC01) 6763231H1 (BRAUNOR01) 5729256H1 (UTRSTUT05) 5652453H1 (COLNNOT27) 7650036J2 (STOMTDE01)	2245 3132 2870 2376 1068 1 1787 1682 3537 577	2825 3698 3429 2925 1735 706 2332 2328 4149 1323
34	378504CB1	4149	1-578, 1075-2335, 2896-2916, 3393-3434, 4121-4149			

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
35	5275371CB1	3080	1-924, 2938- 3080	5275371T6 (OVARDIN02)	2490	3080
				7757715H1 (SPLNTUE01)	501	1004
				7165129R8 (PLACNOR01)	1	640
				7757715J1 (SPLNTUE01)	2132	2805
				7761028J1 (THYMNOE02)	1924	2682
				4749754F6 (SMCRUNT01)	1348	1853
				7653655H1 (UTREDME06)	1582	2053
				7755635H1 (SPLNTUE01)	653	1391
				7267089H2 (NOSEDIC01)	1901	2501
				5371126H1 (BRAINOT22)	2858	3187
36	490576CB1	4167	1339-2048, 802-888, 3127-3570	7102538H1 (BRAWTDRO2)	1215	1715
				6949266R8 (BRAITDRO2)	493	1256
				7193988H1 (BRATDICO1)	1582	2189
				6447351H1 (BRAINOC01)	3702	4167
				490576R6 (HMT2AGT01)	2375	2870
				7066507H1 (BRATNOR01)	681	1271
				71719989V1	1	619
				5961185H1 (BRATNOT05)	2553	3086
				673402R6 (CRBLNOT01)	1290	1725
				4058828F6 (BRAINOT21)	3140	3829
				71207642V1	2548	3175
				6851160H1 (BRAIFEN08)	1855	2518
				7632225H1 (BLADTUE01)	1258	1898
				7283241H1 (BMARTXE01)	819	1385
				7697068H1 (KIDPTDE01)	759	1377
				6307274H1 (NERDTDN03)	2704	3183
				8071039J1 (KIDEUNE02)	3301	3591
37	1417657CB1	3591	1-195, 3299- 3591	2930538H1 (TLYMNOT04)	2215	2567
				289833R6 (TMLR3DT01)	2826	3377
				g2329747	1	287
				8120984H1 (TONSDIC01)	118	794
				5020979F6 (OVARNON03)	1520	2024

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
38	1773215CB1	3685	2601-2740, 1751-2026, 886-931	1773215R6 (MENTUNON3)	2642	3224
				6796337H1 (LIVRTXS02)	2496	3192
				7047512H1 (BRACNOK02)	1901	2596
				6535832H1 (OVARIN02)	3076	3685
				70568996V1	1379	1941
				7220062H1 (SPLNDIC01)	1	690
				6756363H1 (SINTFER02)	649	1358
				70569092V1	1444	1969
				70572278V1	2053	2625
				71885641V1	737	1421
39	3036986CB1	3143	943-2168, 1- 418	5218141T6 (BRSTNOT35)	2022	2459
				7654654J1 (UTREDME06)	1338	1997
				8199424J1 (BRAINOR03)	1	417
				8194678H1 (PROSUNR01)	53	869
				3366420F7 (CONNTUT04)	2914	3143
				66705238	2571	2861
				71714139V1	674	1349
				95177590	2847	3133
				3036986H1 (SMCCNOT01)	1825	2089
				5508168F6 (BRADDIR01)	457	1072
40	2041080CB1	1759	1-241, 841- 1107, 1365- 1759	4152886T8 (MUSLTMT01)	2309	2770
				6557187T8 (BRAFNON02)	986	1759
				7586586H1 (BRAIFEC01)	567	1146
				6558136F6 (BRAFNON02)	66	818
				7977907H1 (LSUBDMC01)	1	467

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
21	105283CB1	MCLDXT02
22	3350821CB1	MCLRUNT01
23	5876846CB1	BRNDIT01
24	3560269CB1	BRSTTUT16
25	4596874CB1	UCMCL5T01
26	3594012CB1	PROSBPT07
27	7482435CB1	TESTTUT03
28	3882333CB1	BLADTUT05
29	7482809CB1	DRGTNON04
30	1739178CB1	FIBRTXS07
31	7473630CB1	BRAUNOT01
32	1431520CB1	PROSTUS23
33	1916304CB1	UTRSNOT08
34	378504CB1	PENITUT01
35	5275371CB1	THYMNOR02
36	490576CB1	HNT2AGT01
37	1417657CB1	TONSDIC01
38	1773215CB1	SPLNFET02
39	3036986CB1	MUSLTM01
40	2041080CB1	BRAFNON02

Table 6

Library	Vector	Library Description
BLADTUT05	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder. Patient history included lung neoplasm and tobacco abuse in remission. Family history included malignant breast neoplasm, tuberculosis, arteriosclerosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer.
BRAFNON02	pINCY	This normalized frontal cortex tissue library was constructed from 10.6 million independent clones from a frontal cortex tissue library. Starting RNA was made from superior frontal cortex tissue removed from a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Grossly, the brain regions examined and cranial nerves were unremarkable. No atherosclerosis of the major vessels was noted. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. There were also multiple small microscopic areas of cavitation with surrounding gliosis scattered throughout the cerebral cortex. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly, and an enlarged spleen and liver. Patient medications included simethicone, Lasix, Digoxin, Colace, Zantac, captopril, and Vasotec. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAUNOT01	pINCY	The library was constructed using RNA isolated from caudate/putamen/nucleus accumbens tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomeningeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.
BRONDIT01	pINCY	Library was constructed using RNA isolated from right lower lobe bronchial tissue removed from a pool of 3 asthmatic Caucasian male and female donors, 22- to 51-years-old during bronchial pinch biopsies. Patient history included atopy as determined by positive skin tests to common aero-allergens.
BRSTTUT16	pINCY	Library was constructed using RNA isolated from breast tumor tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated recurrent grade 4, nuclear grade 3, ductal carcinoma. Angiolymphatic space invasion was identified. Left breast needle biopsy indicated grade 4 ductal adenocarcinoma. Paraffin embedded tissue was estrogen positive. Patient history included breast cancer and deficiency anemia. Family history included cervical cancer.

Table 6 (cont.)

Library	Vector	Library Description
DRGTN0N04	pINCY	The normalized dorsal root ganglion tissue library was constructed from 5.64 million independent clones from the a dorsal root ganglion library. Starting RNA was made from thoracic dorsal root ganglion tissue from a 32-year-old Caucasian male, who died from acute pulmonary edema, acute bronchopneumonia, pleural and pericardial effusion, and lymphoma. The patient presented with pyrexia, fatigue, and GI bleeding. Patient history included probable cytomegalovirus infection, liver congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. The library was normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48-hours/round) reannealing hybridization was used. The library was then linearized and recircularized to select for insert containing clones as follows: plasmid DNA was prepped from approximately 1 million clones from the normalized dorsal root ganglion tissue library following soft agar transformation.
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from the an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
HNT2AGT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937233), using RNA isolated from the hNT2 cell line derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor. Cells were treated with retinoic acid for 5 weeks and with mitotic inhibitors for two weeks and allowed to mature for an additional 4 weeks in conditioned medium.
MCLDXT02	pINCY	Library was constructed using RNA isolated from treated umbilical cord blood dendritic cells removed from a male. The cells were treated with granulocyte/macrophage colony stimulating factor (GM-CSF), tumor necrosis factor alpha (TNF alpha), stem cell factor (SCF), phorbol myristate acetate (PMA), and ionomycin. The GM-CSF was added at time 0 at 100 ng/ml, the TNF alpha was added at time 0 at 2.5 ng/ml, the SCF was added at time 0 at 25 ng/ml. The PMA and ionomycin were added at 13 days for five hours. Incubation time was 13 days.
MCLRUNT01	pINCY	The library was constructed using RNA isolated from untreated peripheral blood mononuclear cell tissue obtained from buffy coat and removed from a 60-year-old male.

Table 6 (cont.)

Library	Vector	Library Description
MUSLTMT01	pINCY	Library was constructed using RNA isolated from glossal muscle tissue removed from a 41-year-old Caucasian female during partial glossectomy. Pathology indicated the excision margins were negative for tumor. Pathology for the matched tumor tissue indicated invasive grade 3, squamous cell carcinoma forming an ulcerated mass of the tongue. The patient presented with a complicated open wound of the tongue. Patient history included obesity, an unspecified nasal and sinus disease, and normal delivery. Patient medications included Premarin, Hydrocodone, vitamins, and Equate nasal spray. Family history included benign hypertension, atherosclerotic coronary artery disease, upper lobe lung cancer, type II diabetes, hyperlipidemia, and cirrhosis of the liver in the father.
PENITUT01	pINCY	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
PROSBPT07	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 53-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma (Gleason grade 3+2). The patient presented with elevated prostate specific antigen and induration. Patient history included hyperlipidemia. Family history included atherosclerotic coronary artery disease, coronary artery bypass graft, perforated gallbladder, hyperlipidemia, and kidney stones.
PROSTUS23	pINCY	This subtracted prostate tumor library was constructed using 10 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 10 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenocarcinoma in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Ronaldo, et al. Genome Research 6 (1996):791.
SPLNFET02	pINCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
TESTTUT03	pINCY	Library was constructed using RNA isolated from right testicular tumor tissue removed from a 45-year-old Caucasian male during a unilateral orchiectomy. Pathology indicated seminoma. Patient history included hyperlipidemia and stomach ulcer. Family history included cerebrovascular disease, skin cancer, hyperlipidemia, acute myocardial infarction, and atherosclerotic coronary artery disease.
THYMNOR02	pINCY	The library was constructed using RNA isolated from thymus tissue removed from a 2-year-old Caucasian female during a thymectomy and patch closure of left atrioventricular fistula. Pathology indicated there was no gross abnormality of the thymus. The patient presented with congenital heart abnormalities. Patient history included double inlet left ventricle and a rudimentary right ventricle, pulmonary hypertension, cyanosis, subaortic stenosis, seizures, and a fracture of the skull base. Family history included reflux neuropathy.
TONSDIC01	PSPORT1	This large size fractionated library was constructed using pooled cDNA from two donors. cDNA was generated using mRNA isolated from diseased left tonsil tissue removed from a 6-year-old Caucasian male (donor A) during adenotonsillectomy and from diseased right tonsil tissue removed from a 9-year-old Caucasian female (donor B) during adenotonsillectomy. Pathology indicated reactive lymphoid hyperplasia, bilaterally (A) and lymphoid hyperplasia (B). The patients presented with sleep apnea (A) and hypertrophy of tonsils, cough, and unspecified nasal and sinus disease (B). Patient history included a bacterial infection (A). Previous surgeries included myringotomy with tube insertion (A). Donor A was not taking any medications and donor B was taking Vancenase. Family history included benign hypertension, myocardial infarction, and atherosclerotic coronary artery disease in the grandparent(s) of donor A; and extrinsic asthma and unspecified allergy in the mother; unspecified allergy in the father; benign hypertension, deficiency anemia, osteoarthritis, extrinsic asthma and unspecified allergy in the grandparent(s) of donor B.
UCMCL5T01	PBLUESCRIPT	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates.
UTRSNOT08	pINCY	Library was constructed using RNA isolated from uterine tissue removed from a 35-year-old Caucasian female during a vaginal hysterectomy with dilation and curettage. Pathology indicated that the endometrium was secretory phase with a benign endometrial polyp 1 cm in diameter. The cervix showed mild chronic cervicitis. Family history included atherosclerotic coronary artery disease and type II diabetes.



Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fasta E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score > GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

- 5 a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,
- b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-20,
- 10 c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, and
- d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

15 2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

20 4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40.

25 6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

30 9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide

- encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from  
5 the group consisting of SEQ ID NO:1-20.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- 10 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:21-40,
- 15 c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a  
20 polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if  
30 present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

19. A method for treating a disease or condition associated with decreased expression of functional INTSIG, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional INTSIG, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and

a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional INTSIG, comprising administering to a patient in need of such treatment a composition of claim 24.

5

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 10 b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 15 a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test  
20 compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

25 28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- 30 b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,  
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,  
c) quantifying the amount of hybridization complex, and  
d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of INTSIG in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and  
b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,  
b) a single chain antibody,  
c) a Fab fragment,  
d) a F(ab')<sub>2</sub> fragment, or  
e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of INTSIG in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of INTSIG in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

5 36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- 10 b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

15 37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

20 39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- 25 c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ
- 30 ID NO:1-20.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.



42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 in the sample.

45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-20.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide

or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is  
5 completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

10 51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

15 53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

20

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at  
25 another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

30

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.

75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.
78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.
- 5 79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.
80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.
81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.
- 10 82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.
83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.
- 15 84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.
85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.
86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.
- 20 87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.
88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.
- 25 89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.
90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.
91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.
- 30 92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.
93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.
95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

## &lt;110&gt; INCYTE GENOMICS, INC.

BAUGHN, Mariah R.  
 DING, Li  
 ELLIOTT, Vicki S.  
 GANDHI, Ameena R.  
 Gietzen, Kimberly J.  
 GRIFFIN, Jennifer A.  
 GURURAJAN, Rajagopal  
 HAFALIA, April J.A.  
 KEARNEY, Liam  
 KHAN, Farrah A.  
 LAL, Preeti  
 LEE, Ernestine A.  
 LU, Dyung Aina M.  
 LU, Yan  
 NGUYEN, Danniel B.  
 ARVIZU, Chandra  
 RAMKUMAR, Jaya  
 TANG, Y. Tom  
 THANGAVELU, Kavitha  
 THORNTON, Michael  
 WALIA, Narinder K.  
 WARREN, Bridget A.  
 XU, Yuming  
 YAO, Monique G.  
 YUE, Henry

## &lt;120&gt; INTRACELLULAR SIGNALING MOLECULES

## &lt;130&gt; PF-0827 PCT

## &lt;140&gt; To Be Assigned

## &lt;141&gt; Herewith

&lt;150&gt; 60/240,871; 60/244,723; 60/249,402; 60/252,622; 60/255,622

&lt;151&gt; 2000-10-13; 2000-10-30; 2000-11-15; 2000-11-22; 2000-12-13

## &lt;160&gt; 40

## &lt;170&gt; PERL Program

## &lt;210&gt; 1

## &lt;211&gt; 805

## &lt;212&gt; PRT

## &lt;213&gt; Homo sapiens

## &lt;220&gt;

## &lt;221&gt; misc\_feature

## &lt;223&gt; Incyte ID No: 105283CD1

## &lt;400&gt; 1

Met	Ala	Ala	Ser	Gly	Val	Pro	Arg	Gly	Cys	Asp	Ile	Leu	Ile	Val
1				5				10					15	
Tyr	Ser	Pro	Asp	Ala	Glu	Glu	Trp	Cys	Gln	Tyr	Leu	Gln	Thr	Leu
				20				25					30	
Phe	Leu	Ser	Ser	Arg	Gln	Val	Arg	Ser	Gln	Lys	Ile	Leu	Thr	His
				35				40					45	
Arg	Leu	Gly	Pro	Glu	Ala	Ser	Phe	Ser	Ala	Glu	Asp	Leu	Ser	Leu
				50				55					60	

Phe	Leu	Ser	Thr	Arg	Cys	Val	Val	Val	Leu	Leu	Ser	Ala	Glu	Leu	
				65					70					75	
Val	Gln	His	Phe	His	Lys	Pro	Ala	Leu	Leu	Pro	Leu	Leu	Gln	Arg	
				80					85					90	
Ala	Phe	His	Pro	Pro	His	Arg	Val	Val	Arg	Leu	Leu	Cys	Gly	Val	
				95					100					105	
Arg	Asp	Ser	Glu	Glu	Phe	Leu	Asp	Phe	Phe	Pro	Asp	Trp	Ala	His	
				110					115					120	
Trp	Gln	Glu	Leu	Thr	Cys	Asp	Asp	Glu	Pro	Glu	Thr	Tyr	Val	Ala	
				125					130					135	
Ala	Val	Lys	Lys	Ala	Ile	Ser	Glu	Asp	Ser	Gly	Cys	Asp	Ser	Val	
				140					145					150	
Thr	Asp	Thr	Glu	Pro	Glu	Asp	Glu	Lys	Val	Val	Ser	Tyr	Ser	Lys	
				155					160					165	
Gln	Gln	Asn	Leu	Pro	Thr	Val	Thr	Ser	Pro	Gly	Asn	Leu	Met	Val	
				170					175					180	
Val	Gln	Pro	Asp	Arg	Ile	Arg	Cys	Gly	Ala	Glu	Thr	Thr	Val	Tyr	
				185					190					195	
Val	Ile	Val	Arg	Cys	Lys	Leu	Asp	Asp	Arg	Val	Ala	Thr	Glu	Ala	
				200					205					210	
Glu	Phe	Ser	Pro	Glu	Asp	Ser	Pro	Ser	Val	Arg	Met	Glu	Ala	Lys	
				215					220					225	
Val	Glu	Asn	Glu	Tyr	Thr	Ile	Ser	Val	Lys	Ala	Pro	Asn	Leu	Ser	
				230					235					240	
Ser	Gly	Asn	Val	Ser	Leu	Lys	Ile	Tyr	Ser	Gly	Asp	Leu	Val	Val	
				245					250					255	
Cys	Glu	Thr	Val	Ile	Ser	Tyr	Tyr	Thr	Asp	Met	Glu	Glu	Ile	Gly	
				260					265					270	
Asn	Leu	Leu	Ser	Asn	Ala	Ala	Asn	Pro	Val	Glu	Phe	Met	Cys	Gln	
				275					280					285	
Ala	Phe	Lys	Ile	Val	Pro	Tyr	Asn	Thr	Glu	Thr	Leu	Asp	Lys	Leu	
				290					295					300	
Leu	Thr	Glu	Ser	Leu	Lys	Asn	Asn	Ile	Pro	Ala	Ser	Gly	Leu	His	
				305					310					315	
Leu	Phe	Gly	Ile	Asn	Gln	Leu	Glu	Glu	Glu	Asp	Met	Met	Thr	Asn	
				320					325					330	
Gln	Arg	Asp	Glu	Glu	Leu	Pro	Thr	Leu	Leu	His	Phe	Ala	Ala	Lys	
				335					340					345	
Tyr	Gly	Leu	Lys	Asn	Leu	Thr	Ala	Leu	Leu	Leu	Thr	Cys	Pro	Gly	
				350					355					360	
Ala	Leu	Gln	Ala	Tyr	Ser	Val	Ala	Asn	Lys	His	Gly	His	Tyr	Pro	
				365					370					375	
Asn	Thr	Ile	Ala	Glu	Lys	His	Gly	Phe	Arg	Asp	Leu	Arg	Gln	Phe	
				380					385					390	
Ile	Asp	Glu	Tyr	Val	Glu	Thr	Val	Asp	Met	Leu	Lys	Ser	His	Ile	
				395					400					405	
Lys	Glu	Glu	Leu	Met	His	Gly	Glu	Glu	Ala	Asp	Ala	Val	Tyr	Glu	
				410					415					420	
Ser	Met	Ala	His	Leu	Ser	Thr	Asp	Leu	Leu	Met	Lys	Cys	Ser	Leu	
				425					430					435	
Asn	Pro	Gly	Cys	Asp	Glu	Asp	Leu	Tyr	Glu	Ser	Met	Ala	Ala	Phe	
				440					445					450	
Val	Pro	Ala	Ala	Thr	Glu	Asp	Leu	Tyr	Val	Glu	Met	Leu	Gln	Ala	
				455					460					465	
Ser	Thr	Ser	Asn	Pro	Ile	Pro	Gly	Asp	Gly	Phe	Ser	Arg	Ala	Thr	
				470					475					480	
Lys	Asp	Ser	Met	Ile	Arg	Lys	Phe	Leu	Glu	Gly	Asn	Ser	Met	Gly	
				485					490					495	

Met Thr Asn Leu	Glu Arg Asp Gln Cys	His Leu Gly Gln Glu Glu	500	505	510
Asp Val Tyr His	Thr Val Asp Asp Asp	Glu Ala Phe Ser Val Asp	515	520	525
Leu Ala Ser Arg	Pro Pro Val Pro Val	Pro Arg Pro Glu Thr Thr	530	535	540
Ala Pro Gly Ala	His Gln Leu Pro Asp	Asn Glu Pro Tyr Ile Phe	545	550	555
Lys Val Phe Ala	Glu Lys Ser Gln Glu	Arg Pro Gly Asn Phe Tyr	560	565	570
Val Ser Ser Glu	Ser Ile Arg Lys Gly	Pro Pro Val Arg Pro Trp	575	580	585
Arg Asp Arg Pro	Gln Ser Ser Ile Tyr	Asp Pro Phe Ala Gly Met	590	595	600
Lys Thr Pro Gly	Gln Arg Gln Leu Ile	Thr Leu Gln Glu Gln Val	605	610	615
Lys Leu Gly Ile	Val Asn Val Asp Glu	Ala Val Leu His Phe Lys	620	625	630
Glu Trp Gln Leu	Asn Gln Lys Arg Arg	Ser Glu Ser Phe Arg Phe	635	640	645
Gln Gln Glu Asn	Leu Lys Arg Leu Arg	Asp Ser Ile Thr Arg Arg	650	655	660
Gln Arg Glu Lys	Gln Lys Ser Gly Lys	Gln Thr Asp Leu Glu Ile	665	670	675
Thr Val Pro Ile	Arg His Ser Gln His	Leu Pro Ala Lys Val Glu	680	685	690
Phe Gly Val Tyr	Glu Ser Gly Pro Arg	Lys Ser Val Ile Pro Pro	695	700	705
Arg Thr Glu Leu	Arg Arg Gly Asp Trp	Lys Thr Asp Ser Thr Ser	710	715	720
Ser Thr Ala Ser	Ser Thr Ser Asn Arg	Ser Ser Thr Arg Ser Leu	725	730	735
Leu Ser Val Ser	Ser Gly Met Glu Gly	Asp Asn Glu Asp Asn Glu	740	745	750
Val Pro Glu Val	Thr Arg Ser Arg Ser	Pro Gly Pro Pro Gln Val	755	760	765
Asp Gly Thr Pro	Thr Met Ser Leu Glu	Arg Pro Pro Arg Val Pro	770	775	780
Pro Arg Ala Ala	Ser Gln Arg Pro Pro	Thr Arg Glu Thr Phe His	785	790	795
Pro Pro Pro Pro	Val Pro Pro Arg Gly	Arg	800	805	

&lt;210&gt; 2

&lt;211&gt; 957

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens .

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3350821CD1

&lt;400&gt; 2

Met Cys Cys Thr	Glu Gly Ser Leu Arg	Lys Arg Asp Ser Gln Arg	1	5	10	15
Ala Pro Glu Ala	Val Leu Cys Leu Gln	Leu Trp Gln Arg Thr Val	20	25	30	
Pro Leu Asp Thr	Leu Lys Gly Leu Gly	Thr Cys Phe Pro Ser Gly				



	35		40		45
Pro Glu Leu Arg Gly Ala Gly Ile Ala Ala Met Glu Arg Ala					
	50		55		60
Ser Glu Arg Arg Thr Ala Ser Ala Leu Phe Ala Gly Phe Arg Ala					
	65		70		75
Leu Gly Leu Phe Ser Asn Asp Ile Pro His Val Val Arg Phe Ser					
	80		85		90
Ala Leu Lys Arg Arg Phe Tyr Val Thr Thr Cys Val Gly Lys Ser					
	95		100		105
Phe His Thr Tyr Asp Val Gln Lys Leu Ser Leu Val Ala Val Ser					
	110		115		120
Asn Ser Val Pro Gln Asp Ile Cys Cys Met Ala Ala Asp Gly Arg					
	125		130		135
Leu Val Phe Ala Ala Tyr Gly Asn Val Phe Ser Ala Phe Ala Arg					
	140		145		150
Asn Lys Glu Ile Val His Thr Phe Lys Gly His Lys Ala Glu Ile					
	155		160		165
His Phe Leu Gln Pro Phe Gly Asp His Ile Ile Ser Val Asp Thr					
	170		175		180
Asp Gly Ile Leu Ile Ile Trp His Ile Tyr Ser Glu Glu Glu Tyr					
	185		190		195
Leu Gln Leu Thr Phe Asp Lys Ser Val Phe Lys Ile Ser Ala Ile					
	200		205		210
Leu His Pro Ser Thr Tyr Leu Asn Lys Ile Leu Leu Gly Ser Glu					
	215		220		225
Gln Gly Ser Leu Gln Leu Trp Asn Val Lys Ser Asn Gln Lys Tyr					
	230		235		240
Pro Ile Arg Gln Thr Phe Ile Pro Ala Gly Tyr Leu Leu Gly Ala					
	245		250		255
His Gly Leu Lys Thr Gln Ala Pro Ala Val Asp Val Val Ala Ile					
	260		265		270
Gly Leu Met Ser Gly Gln Val Ile Ile His Asn Ile Lys Phe Asn					
	275		280		285
Glu Thr Leu Met Lys Phe Arg Gln Asp Trp Gly Pro Ile Thr Ser					
	290		295		300
Ile Ser Phe Arg Thr Asp Gly His Pro Val Met Ala Ala Gly Ser					
	305		310		315
Pro Cys Gly His Ile Gly Leu Trp Asp Leu Glu Asp Lys Lys Leu					
	320		325		330
Ile Asn Gln Met Arg Asn Ala His Ser Thr Ala Ile Ala Gly Leu					
	335		340		345
Thr Phe Leu His Arg Glu Pro Leu Leu Val Thr Asn Gly Ala Asp					
	350		355		360
Asn Ala Leu Arg Ile Trp Ile Phe Asp Gly Pro Thr Gly Glu Gly					
	365		370		375
Arg Leu Leu Arg Phe Arg Met Gly His Ser Ala Pro Leu Thr Asn					
	380		385		390
Ile Arg Tyr Tyr Gly Gln Asn Gly Gln Gln Ile Leu Ser Ala Ser					
	395		400		405
Gln Asp Gly Thr Leu Gln Ser Phe Ser Thr Val His Glu Lys Phe					
	410		415		420
Asn Lys Ser Leu Gly His Gly Leu Ile Asn Lys Lys Arg Val Lys					
	425		430		435
Arg Lys Gly Leu Gln Asn Thr Met Ser Val Arg Leu Pro Pro Ile					
	440		445		450
Thr Lys Phe Ala Ala Glu Glu Ala Arg Glu Ser Asp Trp Asp Gly					
	455		460		465
Ile Ile Ala Cys His Gln Gly Lys Leu Ser Cys Ser Thr Trp Asn					

	470		475		480
Tyr Gln Lys Ser Thr Ile Gly Ala Tyr Phe Leu Lys Pro Lys Glu					
	485		490		495
Leu Lys Lys Asp Asp Ile Thr Ala Thr Ala Val Asp Ile Thr Ser					
	500		505		510
Cys Gly Asn Phe Ala Val Ile Gly Leu Ser Ser Gly Thr Val Asp					
	515		520		525
Val Tyr Asn Met Gln Ser Gly Ile His Arg Gly Ser Phe Gly Lys					
	530		535		540
Asp Gln Ala His Lys Gly Ser Val Arg Gly Val Ala Val Asp Gly					
	545		550		555
Leu Asn Gln Leu Thr Val Thr Thr Gly Ser Glu Gly Leu Leu Lys					
	560		565		570
Phe Trp Asn Phe Lys Asn Lys Ile Leu Ile His Ser Val Ser Leu					
	575		580		585
Ser Ser Ser Pro Asn Ile Met Leu Leu His Arg Asp Ser Gly Ile					
	590		595		600
Leu Gly Leu Ala Leu Asp Asp Phe Ser Ile Ser Val Leu Asp Ile					
	605		610		615
Glu Ala Arg Lys Ile Val Arg Glu Phe Ser Gly His Gln Gly Gln					
	620		625		630
Ile Asn Asp Met Ala Phe Ser Pro Asp Gly Arg Trp Leu Ile Ser					
	635		640		645
Ala Ala Met Asp Cys Ser Ile Arg Thr Trp Asp Leu Pro Ser Gly					
	650		655		660
Cys Leu Ile Asp Cys Phe Leu Leu Asp Ser Ala Pro Leu Asn Val					
	665		670		675
Ser Met Ser Pro Thr Gly Asp Phe Leu Ala Thr Ser His Val Asp					
	680		685		690
His Leu Gly Ile Tyr Leu Trp Ser Asn Ile Ser Leu Tyr Ser Val					
	695		700		705
Val Ser Leu Arg Pro Leu Pro Ala Asp Tyr Val Pro Ser Ile Val					
	710		715		720
Met Leu Pro Gly Thr Cys Gln Thr Gln Asp Val Glu Val Ser Glu					
	725		730		735
Glu Thr Val Glu Pro Ser Asp Glu Leu Ile Glu Tyr Asp Ser Pro					
	740		745		750
Glu Gln Leu Asn Glu Gln Leu Val Thr Leu Ser Leu Leu Pro Glu					
	755		760		765
Ser Arg Trp Lys Asn Leu Leu Asn Leu Asp Val Ile Lys Lys Lys					
	770		775		780
Asn Lys Pro Lys Glu Pro Pro Lys Val Pro Lys Ser Ala Pro Phe					
	785		790		795
Phe Ile Pro Thr Ile Pro Gly Leu Val Pro Arg Tyr Ala Ala Pro					
	800		805		810
Glu Gln Asn Asn Asp Pro Gln Gln Ser Lys Val Val Asn Leu Gly					
	815		820		825
Val Leu Ala Gln Lys Ser Asp Phe Cys Leu Lys Leu Glu Glu Gly					
	830		835		840
Leu Val Asn Asn Lys Tyr Asp Thr Ala Leu Asn Leu Leu Lys Glu					
	845		850		855
Ser Gly Pro Ser Gly Ile Glu Thr Glu Leu Arg Ser Leu Ser Pro					
	860		865		870
Asp Cys Gly Gly Ser Ile Glu Val Met Gln Ser Phe Leu Lys Met					
	875		880		885
Ile Gly Met Met Leu Asp Arg Lys Arg Asp Phe Glu Leu Ala Gln					
	890		895		900
Ala Tyr Leu Ala Leu Phe Leu Lys Leu His Leu Lys Met Leu Pro					

	905		910		915
Ser Glu Pro Val	Leu Leu Glu Glu Ile	Thr Asn Leu Ser Ser	Gln		
	920		925		930
Val Glu Glu Asn Trp	Thr His Leu Gln	Ser Leu Phe Asn Gln	Ser		
	935		940		945
Met Cys Ile Leu Asn	Tyr Leu Lys Ser	Ala Leu Leu			
	950		955		

&lt;210&gt; 3

&lt;211&gt; 274

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5876846CD1

&lt;400&gt; 3

Met Val Arg Leu Thr	Leu Asp Leu Ile	Ala Arg Asn Ser Asn	Leu
1	5	10	15
Lys Pro Arg Lys Glu	Thr Ile Ser Gln	Cys Leu Lys Lys	Ile
	20	25	30
Thr His Ile Asn Phe	Ser Asp Lys Asn	Ile Asp Ala Ile	Glu Asp
	35	40	45
Leu Ser Leu Cys Lys	Asn Leu Ser Val	Leu Tyr Leu Tyr	Asp Asn
	50	55	60
Cys Ile Ser Gln Ile	Thr Asn Leu Asn	Tyr Ala Thr Asn	Leu Thr
	65	70	75
His Leu Tyr Leu Gln	Asn Asn Cys Ile	Ser Cys Ile Glu	Asn Leu
	80	85	90
Arg Ser Leu Lys Lys	Leu Glu Lys Leu	Tyr Leu Gly Gly	Asn Tyr
	95	100	105
Ile Ala Val Ile Glu	Gly Leu Glu Gly	Leu Gly Glu Leu	Arg Glu
	110	115	120
Leu His Val Glu Asn	Gln Arg Leu Pro	Leu Gly Glu Lys	Leu Leu
	125	130	135
Phe Asp Pro Arg Thr	Leu His Ser Leu	Ala Lys Ser Leu	Cys Ile
	140	145	150
Leu Asn Ile Ser Asn	Asn Asn Ile Asp	Asp Ile Thr Asp	Leu Glu
	155	160	165
Leu Leu Glu Asn Leu	Asn Gln Leu Ile	Ala Val Asp Asn	Gln Leu
	170	175	180
Leu His Val Lys Asp	Leu Glu Phe Leu	Leu Asn Lys Leu	Met Lys
	185	190	195
Leu Trp Lys Ile Asp	Leu Asn Gly Asn	Pro Val Cys Leu	Lys Pro
	200	205	210
Lys Tyr Arg Asp Arg	Leu Ile Leu Val	Ser Lys Ser Leu	Glu Phe
	215	220	225
Leu Asp Gly Lys Glu	Ile Lys Asn Ile	Glu Arg Gln Phe	Leu Met
	230	235	240
Asn Trp Lys Ala Ser	Lys Asp Ala Lys	Lys Ile Ser Lys	Lys Arg
	245	250	255
Ser Ser Lys Asn Glu	Asp Ala Ser Asn	Ser Leu Ile Ser	Lys His
	260	265	270
Ser Val Thr His			

&lt;210&gt; 4

&lt;211&gt; 1144

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3560269CD1

&lt;400&gt; 4

```

Met Pro Thr Ala Glu Ser Glu Ala Lys Val Lys Thr Lys Val Arg
 1          5          10          15
Phe Glu Glu Leu Leu Lys Thr His Ser Asp Leu Met Arg Glu Lys
          20          25          30
Lys Lys Leu Lys Lys Lys Leu Val Arg Ser Glu Glu Asn Ile Ser
          35          40          45
Pro Asp Thr Ile Arg Ser Asn Leu His Tyr Met Lys Glu Thr Thr
          50          55          60
Ser Asp Asp Pro Asp Thr Ile Arg Ser Asn Leu Pro His Ile Lys
          65          70          75
Glu Thr Thr Ser Asp Asp Val Ser Ala Ala Asn Thr Asn Asn Leu
          80          85          90
Lys Lys Ser Thr Arg Val Thr Lys Asn Lys Leu Arg Asn Thr Gln
          95          100          105
Leu Ala Thr Glu Asn Pro Asn Gly Asp Ala Ser Val Glu Glu Asp
          110          115          120
Lys Gln Gly Lys Pro Asn Lys Lys Val Ile Lys Thr Val Pro Gln
          125          130          135
Leu Thr Thr Gln Asp Leu Lys Pro Glu Thr Pro Glu Asn Lys Val
          140          145          150
Asp Ser Thr His Gln Lys Thr His Thr Lys Pro Gln Pro Gly Val
          155          160          165
Asp His Gln Lys Ser Glu Lys Ala Asn Glu Gly Arg Glu Glu Thr
          170          175          180
Asp Leu Glu Glu Asp Glu Glu Leu Met Gln Ala Tyr Gln Cys His
          185          190          195
Val Thr Glu Glu Met Ala Lys Glu Ile Lys Arg Lys Ile Arg Lys
          200          205          210
Lys Leu Lys Glu Gln Leu Thr Tyr Phe Pro Ser Asp Thr Leu Phe
          215          220          225
His Asp Asp Lys Leu Ser Ser Glu Lys Arg Lys Lys Lys Lys Glu
          230          235          240
Val Pro Val Phe Ser Lys Ala Glu Thr Ser Thr Leu Thr Ile Ser
          245          250          255
Gly Asp Thr Val Glu Gly Glu Gln Lys Lys Glu Ser Ser Val Arg
          260          265          270
Ser Val Ser Ser Asp Ser His Gln Asp Asp Glu Ile Ser Ser Met
          275          280          285
Glu Gln Ser Thr Glu Asp Ser Met Gln Asp Asp Thr Lys Pro Lys
          290          295          300
Pro Lys Lys Thr Lys Lys Lys Thr Lys Ala Val Ala Asp Asn Asn
          305          310          315
Glu Asp Val Asp Gly Asp Gly Val His Glu Ile Thr Ser Arg Asp
          320          325          330
Ser Pro Val Tyr Pro Lys Cys Leu Leu Asp Asp Asp Leu Val Leu
          335          340          345
Gly Val Tyr Ile His Arg Thr Asp Arg Leu Lys Ser Asp Phe Met
          350          355          360
Ile Ser His Pro Met Val Lys Ile His Val Val Asp Glu His Thr

```

	365		370		375
Gly Gln Tyr Val Lys Lys Asp Asp Ser Gly Arg Pro Val Ser Ser					
	380		385		390
Tyr Tyr Glu Lys Glu Asn Val Asp Tyr Ile Leu Pro Ile Met Thr					
	395		400		405
Gln Pro Tyr Asp Phe Lys Gln Leu Lys Ser Arg Leu Pro Glu Trp					
	410		415		420
Glu Glu Gln Ile Val Phe Asn Glu Asn Phe Pro Tyr Leu Leu Arg					
	425		430		435
Gly Ser Asp Glu Ser Pro Lys Val Ile Leu Phe Phe Glu Ile Leu					
	440		445		450
Asp Phe Leu Ser Val Asp Glu Ile Lys Asn Asn Ser Glu Val Gln					
	455		460		465
Asn Gln Glu Cys Gly Phe Arg Lys Ile Ala Trp Ala Phe Leu Lys					
	470		475		480
Leu Leu Gly Ala Asn Gly Asn Ala Asn Ile Asn Ser Lys Leu Arg					
	485		490		495
Leu Gln Leu Tyr Tyr Pro Pro Thr Lys Pro Arg Ser Pro Leu Ser					
	500		505		510
Val Val Glu Ala Phe Glu Trp Trp Ser Lys Cys Pro Arg Asn His					
	515		520		525
Tyr Pro Ser Thr Leu Tyr Val Thr Val Arg Gly Leu Lys Val Pro					
	530		535		540
Asp Cys Ile Lys Pro Ser Tyr Arg Ser Met Met Ala Leu Gln Glu					
	545		550		555
Glu Lys Gly Lys Pro Val His Cys Glu Arg His His Glu Ser Ser					
	560		565		570
Ser Val Asp Thr Glu Pro Gly Leu Glu Glu Ser Lys Glu Val Ile					
	575		580		585
Lys Trp Lys Arg Leu Pro Gly Gln Ala Cys Arg Ile Pro Asn Lys					
	590		595		600
His Leu Phe Ser Leu Asn Ala Gly Glu Arg Gly Cys Phe Cys Leu					
	605		610		615
Asp Phe Ser His Asn Gly Arg Ile Leu Ala Ala Ala Cys Ala Ser					
	620		625		630
Arg Asp Gly Tyr Pro Ile Ile Leu Tyr Glu Ile Pro Ser Gly Arg					
	635		640		645
Phe Met Arg Glu Leu Cys Gly His Leu Asn Ile Ile Tyr Asp Leu					
	650		655		660
Ser Trp Ser Lys Asp Asp His Tyr Ile Leu Thr Ser Ser Ser Asp					
	665		670		675
Gly Thr Ala Arg Ile Trp Lys Asn Glu Ile Asn Asn Thr Asn Thr					
	680		685		690
Phe Arg Val Leu Pro His Pro Ser Phe Val Tyr Thr Ala Lys Phe					
	695		700		705
His Pro Ala Val Arg Glu Leu Val Val Thr Gly Cys Tyr Asp Ser					
	710		715		720
Met Ile Arg Ile Trp Lys Val Glu Met Arg Glu Asp Ser Ala Ile					
	725		730		735
Leu Val Arg Gln Phe Asp Val His Lys Ser Phe Ile Asn Ser Leu					
	740		745		750
Cys Phe Asp Thr Glu Gly His His Met Tyr Ser Gly Asp Cys Thr					
	755		760		765
Gly Val Ile Val Val Trp Asn Thr Tyr Val Lys Ile Asn Asp Leu					
	770		775		780
Glu His Ser Val His His Trp Thr Ile Asn Lys Glu Ile Lys Glu					
	785		790		795
Thr Glu Phe Lys Gly Ile Pro Ile Ser Tyr Leu Glu Ile His Pro					

Asn Gly Lys Arg	800	805	810
Leu Leu Ile His Thr		Lys Asp Ser Thr Leu Arg	
815		820	825
Ile Met Asp Leu Arg		Arg Lys Phe Val Gly Ala	
830		835	840
Ala Asn Tyr Arg		Thr Leu Thr Pro Cys Gly	
845		850	855
Thr Phe Leu Phe		Ala Gly Ser Glu Asp Gly Ile Val Tyr Val Trp	
860		865	870
Asn Pro Glu Thr		Gly Glu Gln Val Ala Met Tyr Ser Asp Leu Pro	
875		880	885
Phe Lys Ser Pro		Ile Arg Asp Ile Ser Tyr His Pro Phe Glu Asn	
890		895	900
Met Val Ala Phe		Cys Ala Phe Gly Gln Asn Glu Pro Ile Leu Leu	
905		910	915
Tyr Ile Tyr Asp		Phe His Val Ala Gln Gln Glu Ala Glu Met Phe	
920		925	930
Lys Arg Tyr Asn		Gly Thr Phe Pro Leu Pro Gly Ile His Gln Ser	
935		940	945
Gln Asp Ala Leu		Cys Thr Cys Pro Lys Leu Pro His Gln Gly Ser	
950		955	960
Phe Gln Ile Asp		Glu Phe Val His Thr Glu Ser Ser Ser Thr Lys	
965		970	975
Met Gln Leu Val		Lys Gln Arg Leu Glu Thr Val Thr Glu Val Ile	
980		985	990
Arg Ser Cys Ala		Ala Lys Val Asn Lys Asn Leu Ser Phe Thr Ser	
995		1000	1005
Pro Pro Ala Val		Ser Ser Gln Gln Ser Lys Leu Lys Gln Ser Asn	
1010		1015	1020
Met Leu Thr Ala		Gln Glu Ile Leu His Gln Phe Gly Phe Thr Gln	
1025		1030	1035
Thr Gly Ile Ile		Ser Ile Glu Arg Lys Pro Cys Asn His Gln Val	
1040		1045	1050
Asp Thr Ala Pro		Thr Val Val Ala Leu Tyr Asp Tyr Thr Ala Asn	
1055		1060	1065
Arg Ser Asp Glu		Leu Thr Ile His Arg Gly Asp Ile Ile Arg Val	
1070		1075	1080
Phe Phe Lys Asp		Asn Glu Asp Trp Trp Tyr Gly Ser Ile Gly Lys	
1085		1090	1095
Gly Gln Glu Gly		Tyr Phe Pro Ala Asn His Val Ala Ser Glu Thr	
1100		1105	1110
Leu Tyr Gln Glu		Leu Pro Pro Glu Ile Lys Glu Arg Ser Pro Pro	
1115		1120	1125
Leu Ser Pro Glu		Glu Lys Thr Lys Ile Glu Lys Ser Pro Ala Pro	
1130		1135	1140
Gln Lys Val Lys			

&lt;210&gt; 5

&lt;211&gt; 513

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 4596874CD1

&lt;400&gt; 5

Met	Ala	Tyr	Gln	Val	Val	Glu	Lys	Gly	Ala	Ala	Leu	Gly	Thr	Leu
1				5					10					15
Glu	Ser	Glu	Leu	Gln	Gln	Arg	Gln	Ser	Arg	Leu	Ala	Ala	Leu	Glu
				20					25					30
Ala	Arg	Val	Ala	Gln	Leu	Arg	Glu	Ala	Arg	Ala	Gln	Gln	Ala	Gln
				35					40					45
Gln	Val	Glu	Glu	Trp	Arg	Ala	Gln	Asn	Ala	Val	Gln	Arg	Ala	Ala
				50					55					60
Tyr	Glu	Ala	Leu	Arg	Ala	His	Val	Gly	Leu	Arg	Glu	Ala	Ala	Leu
				65					70					75
Arg	Arg	Leu	Gln	Glu	Glu	Ala	Arg	Asp	Leu	Leu	Glu	Arg	Leu	Val
				80					85					90
Gln	Arg	Lys	Ala	Arg	Ala	Ala	Ala	Glu	Arg	Asn	Leu	Arg	Asn	Glu
				95					100					105
Arg	Arg	Glu	Arg	Ala	Lys	Gln	Ala	Arg	Val	Ser	Gln	Glu	Leu	Lys
				110					115					120
Lys	Ala	Ala	Lys	Arg	Thr	Val	Ser	Ile	Ser	Glu	Gly	Pro	Asp	Thr
				125					130					135
Leu	Gly	Asp	Gly	Met	Arg	Glu	Arg	Arg	Glu	Thr	Leu	Ala	Leu	Ala
				140					145					150
Pro	Glu	Pro	Glu	Pro	Leu	Glu	Lys	Glu	Ala	Cys	Glu	Lys	Trp	Lys
				155					160					165
Arg	Pro	Phe	Arg	Ser	Ala	Ser	Ala	Thr	Ser	Leu	Thr	Leu	Ser	His
				170					175					180
Cys	Val	Asp	Val	Val	Lys	Gly	Leu	Leu	Asp	Phe	Lys	Lys	Arg	Arg
				185					190					195
Gly	His	Ser	Ile	Gly	Gly	Ala	Pro	Glu	Gln	Arg	Tyr	Gln	Ile	Ile
				200					205					210
Pro	Val	Cys	Val	Ala	Ala	Arg	Leu	Pro	Thr	Arg	Ala	Gln	Asp	Val
				215					220					225
Leu	Asp	Ala	His	Leu	Ser	Glu	Val	Asn	Ala	Val	Arg	Phe	Gly	Pro
				230					235					240
Asn	Ser	Ser	Leu	Leu	Ala	Thr	Gly	Gly	Ala	Asp	Arg	Leu	Ile	His
				245					250					255
Leu	Trp	Asn	Val	Val	Gly	Ser	Arg	Leu	Glu	Ala	Asn	Gln	Thr	Leu
				260					265					270
Glu	Gly	Ala	Gly	Gly	Ser	Ile	Thr	Ser	Val	Asp	Phe	Asp	Pro	Ser
				275					280					285
Gly	Tyr	Gln	Val	Leu	Ala	Ala	Thr	Tyr	Asn	Gln	Ala	Ala	Gln	Leu
				290					295					300
Trp	Lys	Val	Gly	Glu	Ala	Gln	Ser	Lys	Glu	Thr	Leu	Ser	Gly	His
				305					310					315
Lys	Asp	Lys	Val	Thr	Ala	Ala	Lys	Phe	Lys	Leu	Thr	Arg	His	Gln
				320					325					330
Ala	Val	Thr	Gly	Ser	Arg	Asp	Arg	Thr	Val	Lys	Glu	Trp	Asp	Leu
				335					340					345
Gly	Arg	Ala	Tyr	Cys	Ser	Arg	Thr	Ile	Asn	Val	Leu	Ser	Tyr	Cys
				350					355					360
Asn	Asp	Val	Val	Cys	Gly	Asp	His	Ile	Ile	Ile	Ser	Gly	His	Asn
				365					370					375
Asp	Gln	Lys	Ile	Arg	Phe	Trp	Asp	Ser	Arg	Gly	Pro	His	Cys	Thr
				380					385					390
Gln	Val	Ile	Pro	Val	Gln	Gly	Arg	Val	Thr	Ser	Leu	Ser	Leu	Ser
				395					400					405
His	Asp	Gln	Leu	His	Leu	Leu	Ser	Cys	Ser	Arg	Asp	Asn	Thr	Leu
				410					415					420
Lys	Val	Ile	Asp	Leu	Arg	Val	Ser	Asn	Ile	Arg	Gln	Val	Phe	Arg
				425					430					435

Ala	Asp	Gly	Phe	Lys	Cys	Gly	Ser	Asp	Trp	Thr	Lys	Ala	Val	Phe
				440					445					450
Ser	Pro	Asp	Arg	Ser	Tyr	Ala	Leu	Ala	Gly	Ser	Cys	Asp	Gly	Ala
				455					460					465
Leu	Tyr	Ile	Trp	Asp	Val	Asp	Thr	Gly	Lys	Leu	Glu	Ser	Arg	Leu
				470					475					480
Gln	Gly	Pro	His	Cys	Ala	Ala	Val	Asn	Ala	Val	Ala	Trp	Cys	Tyr
				485					490					495
Ser	Gly	Ser	His	Met	Val	Ser	Val	Asp	Gln	Gly	Arg	Lys	Val	Val
				500					505					510
Leu	Trp	Gln												

&lt;210&gt; 6

&lt;211&gt; 667

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3594012CD1

&lt;400&gt; 6

Met	His	Phe	Thr	Gly	Ala	Val	Met	Gln	Glu	Asn	Leu	Arg	Phe	Ala
1				5					10					15
Ser	Ser	Gly	Asp	Asp	Ile	Lys	Ile	Trp	Asp	Ala	Ser	Ser	Met	Thr
				20					25					30
Leu	Val	Asp	Lys	Phe	Asn	Pro	His	Thr	Ser	Pro	His	Gly	Ile	Ser
				35					40					45
Ser	Ile	Cys	Trp	Ser	Ser	Asn	Asn	Asn	Phe	Leu	Val	Thr	Ala	Ser
				50					55					60
Ser	Ser	Gly	Asp	Lys	Ile	Val	Val	Ser	Ser	Cys	Lys	Cys	Lys	Pro
				65					70					75
Val	Pro	Leu	Leu	Glu	Leu	Ala	Glu	Gly	Gln	Lys	Gln	Thr	Cys	Val
				80					85					90
Asn	Leu	Asn	Ser	Thr	Ser	Met	Tyr	Leu	Val	Ser	Gly	Gly	Leu	Asn
				95					100					105
Asn	Thr	Val	Asn	Ile	Trp	Asp	Leu	Lys	Ser	Lys	Arg	Val	His	Arg
				110					115					120
Ser	Leu	Lys	Asp	His	Lys	Asp	Gln	Val	Thr	Cys	Val	Thr	Tyr	Asn
				125					130					135
Trp	Asn	Asp	Cys	Tyr	Ile	Ala	Ser	Gly	Ser	Leu	Ser	Gly	Glu	Ile
				140					145					150
Ile	Leu	His	Ser	Val	Thr	Thr	Asn	Leu	Ser	Ser	Thr	Pro	Phe	Gly
				155					160					165
His	Gly	Ser	Asn	Gln	Ser	Val	Arg	His	Leu	Lys	Tyr	Ser	Leu	Phe
				170					175					180
Lys	Lys	Ser	Leu	Leu	Gly	Ser	Val	Ser	Asp	Asn	Gly	Ile	Val	Thr
				185					190					195
Leu	Trp	Asp	Val	Asn	Ser	Gln	Ser	Pro	Tyr	His	Asn	Phe	Asp	Ser
				200					205					210
Val	His	Lys	Ala	Pro	Ala	Ser	Gly	Ile	Cys	Phe	Ser	Pro	Val	Asn
				215					220					225
Glu	Leu	Leu	Phe	Val	Thr	Ile	Gly	Leu	Asp	Lys	Arg	Ile	Ile	Leu
				230					235					240
Tyr	Asp	Thr	Ser	Ser	Lys	Lys	Leu	Val	Lys	Thr	Leu	Val	Ala	Asp
				245					250					255
Thr	Pro	Leu	Thr	Ala	Val	Asp	Phe	Met	Pro	Asp	Gly	Ala	Thr	Leu



Ala Ile Gly Ser	260	265	270
Ser Arg Gly Lys Ile Tyr Gln Tyr Asp Leu Arg			
275	280	285	
Met Leu Lys Ser Pro Val Lys Thr Ile Ser Ala His Lys Thr Ser			
290	295	300	
Val Gln Cys Ile Ala Phe Gln Tyr Ser Thr Val Leu Thr Lys Ser			
305	310	315	
Ser Leu Asn Lys Gly Cys Ser Asn Lys Pro Thr Thr Val Asn Lys			
320	325	330	
Arg Ser Val Asn Val Asn Ala Ala Ser Gly Gly Val Gln Asn Ser			
335	340	345	
Gly Ile Val Arg Glu Ala Pro Ala Thr Ser Ile Ala Thr Val Leu			
350	355	360	
Pro Gln Pro Met Thr Ser Ala Met Gly Lys Gly Thr Val Ala Val			
365	370	375	
Gln Glu Lys Ala Gly Leu Pro Arg Ser Ile Asn Thr Asp Thr Leu			
380	385	390	
Ser Lys Glu Thr Asp Ser Gly Lys Asn Gln Asp Phe Ser Ser Phe			
395	400	405	
Asp Asp Thr Gly Lys Ser Ser Leu Gly Asp Met Phe Ser Pro Ile			
410	415	420	
Arg Asp Asp Ala Val Val Asn Lys Gly Ser Asp Glu Ser Ile Gly			
425	430	435	
Lys Gly Asp Gly Phe Asp Phe Leu Pro Gln Leu Asn Ser Val Phe			
440	445	450	
Pro Pro Arg Lys Asn Pro Val Thr Ser Ser Thr Ser Val Leu His			
455	460	465	
Ser Ser Pro Leu Asn Val Phe Met Gly Ser Pro Gly Lys Glu Glu			
470	475	480	
Asn Glu Asn Arg Asp Leu Thr Ala Glu Ser Lys Lys Ile Tyr Met			
485	490	495	
Gly Lys Gln Glu Ser Lys Asp Ser Phe Lys Gln Leu Ala Lys Leu			
500	505	510	
Val Thr Ser Gly Ala Glu Ser Gly Asn Leu Asn Thr Ser Pro Ser			
515	520	525	
Ser Asn Gln Thr Arg Asn Ser Glu Lys Phe Glu Lys Pro Glu Asn			
530	535	540	
Glu Ile Glu Ala Gln Leu Ile Cys Glu Pro Pro Ile Asn Gly Ser			
545	550	555	
Ser Thr Pro Asn Pro Lys Ile Ala Ser Ser Val Thr Ala Gly Val			
560	565	570	
Ala Ser Ser Leu Ser Glu Lys Ile Ala Asp Ser Ile Gly Asn Asn			
575	580	585	
Arg Gln Asn Ala Pro Leu Thr Ser Ile Gln Ile Arg Phe Ile Gln			
590	595	600	
Asn Met Ile Gln Glu Thr Leu Asp Asp Phe Arg Glu Ala Cys His			
605	610	615	
Arg Asp Ile Val Asn Leu Gln Val Glu Met Ile Lys Gln Phe His			
620	625	630	
Met Gln Leu Asn Glu Met His Ser Leu Leu Glu Arg Tyr Ser Val			
635	640	645	
Asn Glu Gly Leu Val Ala Glu Ile Glu Arg Leu Arg Glu Glu Asn			
650	655	660	
Lys Arg Leu Arg Ala His Phe			
665			

&lt;210&gt; 7

&lt;211&gt; 897

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7482435CD1

&lt;400&gt; 7

Met	Ala	Glu	Ile	Pro	Leu	Tyr	Phe	Val	Asp	Leu	Gln	Asp	Asp	Leu	1	5	10	15
Asp	Asp	Tyr	Gly	Phe	Glu	Asp	Tyr	Gly	Thr	Asp	Cys	Asp	Asn	Met	20	25	30	
Arg	Val	Thr	Ala	Phe	Leu	Asp	Ile	Pro	Gly	Gln	Asp	Asn	Leu	Pro	35	40	45	
Pro	Leu	Thr	Arg	Leu	Glu	Lys	Tyr	Ala	Phe	Ser	Glu	Asn	Thr	Phe	50	55	60	
Asn	Arg	Gln	Ile	Ile	Ala	Arg	Gly	Leu	Leu	Asp	Ile	Phe	Arg	Asp	65	70	75	
Phe	Gly	Asn	Asn	Glu	Glu	Asp	Phe	Leu	Thr	Val	Met	Glu	Ile	Val	80	85	90	
Val	Arg	Leu	Ser	Glu	Asp	Ala	Glu	Pro	Thr	Val	Arg	Thr	Glu	Leu	95	100	105	
Met	Glu	Gln	Ile	Pro	Pro	Ile	Ala	Ile	Phe	Leu	Gln	Glu	Asn	Arg	110	115	120	
Ser	Asn	Phe	Pro	Val	Val	Leu	Ser	Glu	Tyr	Leu	Ile	Pro	Ile	Val	125	130	135	
Met	Arg	Tyr	Leu	Thr	Asp	Pro	Asn	Asn	Gln	Ile	Ile	Cys	Lys	Met	140	145	150	
Ala	Ser	Met	Leu	Ser	Lys	Ser	Thr	Val	Glu	Arg	Leu	Leu	Leu	Pro	155	160	165	
Arg	Phe	Cys	Glu	Leu	Cys	Gly	Asp	Arg	Lys	Leu	Phe	Gln	Val	Arg	170	175	180	
Lys	Val	Cys	Ala	Ala	Asn	Phe	Gly	Asp	Ile	Cys	His	Ala	Val	Gly	185	190	195	
Gln	Glu	Ala	Thr	Glu	Lys	Phe	Leu	Ile	Pro	Lys	Phe	Phe	Glu	Leu	200	205	210	
Cys	Ser	Asp	Ala	Val	Trp	Gly	Met	Arg	Lys	Ala	Cys	Ala	Glu	Cys	215	220	225	
Phe	Thr	Ala	Val	Ser	His	Ser	Ser	Ser	Pro	Gly	Val	Arg	Arg	Thr	230	235	240	
Gln	Leu	Phe	Pro	Leu	Phe	Ile	Arg	Leu	Val	Ser	Asp	Pro	Cys	Arg	245	250	255	
Trp	Val	His	Gln	Ala	Ala	Phe	Gln	Ser	Leu	Gly	Pro	Phe	Ile	Ser	260	265	270	
Thr	Phe	Ala	Asn	Pro	Ser	Arg	Ala	Gly	Leu	Tyr	Leu	Arg	Glu	Asp	275	280	285	
Gly	Ala	Leu	Ser	Ile	Trp	Pro	Leu	Thr	Gln	Asp	Leu	Asp	Ser	Gly	290	295	300	
Phe	Ala	Ser	Gly	Ser	Pro	Ala	Pro	Ser	Ser	Gly	Gly	Asn	Ile	Ser	305	310	315	
Pro	Ala	Ser	Leu	Ile	Arg	Ser	Ala	Lys	Pro	Val	Arg	Ser	Glu	Pro	320	325	330	
Glu	Leu	Pro	Val	Glu	Gly	Thr	Ser	Ala	Lys	Thr	Ser	Asp	Cys	Pro	335	340	345	
His	Ser	Ser	Ser	Ser	Ser	Asp	Gly	Pro	Ala	Glu	Ser	Pro	Val	Glu	350	355	360	
Ser	Cys	Val	Ser	Ala	Gly	Ala	Glu	Trp	Thr	Arg	Val	Ser	Pro	Glu	365	370	375	

Thr Ser Ala Arg	Ser Lys Leu Ser Asp	Met Asn Asp Leu Pro Ile	
	380	385	390
Ser Ser Tyr Pro	Gly Ser Asp Ser Trp	Ala Cys Pro Gly Asn Thr	
	395	400	405
Glu Asp Val Phe	Ser His Phe Leu Tyr	Cys Lys Asp Leu Glu Leu	
	410	415	420
Leu Leu Ser Glu	Ala Gly Pro Gln Glu	Asp Asp Cys Ser Arg Pro	
	425	430	435
Gly Val Val His	Asn Ser Cys Val Ala	Arg Ser Glu Ile Gln Lys	
	440	445	450
Val Leu Asp Ser	Leu Gln Glu His Leu	Met Asn Asp Pro Asp Val	
	455	460	465
Gln Ala Gln Val	Gln Val Leu Ser Ala	Ala Leu Arg Ala Ala Gln	
	470	475	480
Leu Asp Cys Val	Asn Glu Ala Glu Ser	Lys Pro Thr Ala Gly Leu	
	485	490	495
Lys Glu Val Ser	Ile Ser His Pro Ser	Ser Ala Ser Asp Asn Gln	
	500	505	510
Ile Ala Leu Ala	Ala Ser Ser Ser Gln	Asp Glu Leu Phe Val Ala	
	515	520	525
Arg Ile Leu Gln	Ser Pro Asp Pro Gly	Gly Pro Arg Asn Gly Thr	
	530	535	540
Ser Asp His Leu	Glu Thr Asp Gln Arg	Gln Asp Pro Thr Pro Leu	
	545	550	555
Glu Glu Asn Lys	Ser Lys Leu Gln Asp	Val Ile Pro Gln Pro Leu	
	560	565	570
Leu Asp Gln Tyr	Val Ser Met Thr Asp	Pro Ala Arg Ala Gln Thr	
	575	580	585
Val Asp Thr Asp	Ile Ala Lys His Cys	Ala Tyr Ser Leu Pro Gly	
	590	595	600
Val Ala Leu Thr	Leu Gly Arg Gln Asn	Trp His Cys Leu Lys Asp	
	605	610	615
Thr Tyr Glu Thr	Leu Ala Ser Asp Val	Gln Trp Lys Val Arg Arg	
	620	625	630
Ala Leu Ala Phe	Ser Ile His Glu Leu	Ala Val Ile Leu Gly Asp	
	635	640	645
Gln Leu Thr Ala	Ala Asp Leu Val Pro	Ile Phe Asn Gly Phe Leu	
	650	655	660
Lys Asp Leu Asp	Glu Val Arg Ile Gly	Val Leu Arg His Leu Tyr	
	665	670	675
Asp Phe Leu Lys	Leu Leu His Glu Asp	Lys Arg Arg Asp Tyr Leu	
	680	685	690
Tyr Gln Leu Gln	Glu Phe Val Val Thr	Asp Asn Ser Arg Asn Trp	
	695	700	705
Arg Phe Arg Tyr	Glu Leu Ala Glu Gln	Leu Ile Leu Ile Leu Glu	
	710	715	720
Leu Tyr Ser Pro	Asn Asp Val Tyr Asp	Tyr Leu Met His Ile Ala	
	725	730	735
Leu Lys Leu Cys	Ala Asp Gln Val Ser	Glu Val Arg Trp Ile Ser	
	740	745	750
Phe Lys Leu Val	Val Ala Ile Leu Gln	Lys Phe Tyr Ser Asn Ser	
	755	760	765
Glu Ser Ala Leu	Gly Leu Asn Phe Ile	Asn Glu Leu Ile Ile Arg	
	770	775	780
Phe Arg His Cys	Ser Lys Trp Val Gly	Arg Gln Ala Phe Ala Phe	
	785	790	795
Ile Cys Gln Ala	Val Val Ser Lys Glu	Cys Val Pro Val Asp Gln	
	800	805	810

Phe Met Glu His Leu Leu Pro Ser Leu Leu Ser Leu Ala Ser Asp  
 815 820 825  
 Pro Val Pro Asn Val Arg Val Leu Leu Ala Lys Ala Leu Arg Gln  
 830 835 840  
 Met Leu Leu Glu Lys Ala Tyr Phe Arg Asn Ala Gly Asn Pro His  
 845 850 855  
 Leu Glu Val Ile Glu Glu Thr Ile Leu Ala Leu Gln Ser Asp Arg  
 860 865 870  
 Asp Gln Asp Val Ser Phe Phe Ala Ala Leu Glu Pro Lys Arg Arg  
 875 880 885  
 Asn Ile Ile Asp Thr Ala Val Leu Glu Lys Gln Asn  
 890 895

<210> 8

<211> 454

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3882333CD1

<400> 8

Met Leu Lys Gln Ile Leu Ser Glu Met Tyr Ile Asp Pro Asp Leu  
 1 5 10 15  
 Leu Ala Glu Leu Ser Glu Glu Gln Lys Gln Ile Leu Phe Phe Lys  
 20 25 30  
 Met Arg Glu Glu Gln Ile Arg Arg Trp Lys Glu Arg Glu Ala Ala  
 35 40 45  
 Met Glu Arg Lys Glu Ser Leu Pro Val Lys Pro Arg Pro Lys Lys  
 50 55 60  
 Glu Asn Gly Lys Ser Val His Trp Lys Leu Gly Ala Asp Lys Glu  
 65 70 75  
 Val Trp Val Trp Val Met Gly Glu His His Leu Asp Lys Pro Tyr  
 80 85 90  
 Asp Val Leu Cys Asn Glu Ile Ile Ala Glu Arg Ala Arg Leu Lys  
 95 100 105  
 Ala Glu Gln Glu Ala Glu Glu Pro Arg Lys Thr His Ser Glu Glu  
 110 115 120  
 Phe Thr Asn Ser Leu Lys Thr Lys Ser Gln Tyr His Asp Leu Gln  
 125 130 135  
 Ala Pro Asp Asn Gln Gln Thr Lys Asp Ile Trp Lys Lys Val Ala  
 140 145 150  
 Glu Lys Glu Glu Leu Glu Gln Gly Ser Arg Pro Ala Pro Thr Leu  
 155 160 165  
 Glu Glu Glu Lys Ile Arg Ser Leu Ser Ser Ser Ser Arg Asn Ile  
 170 175 180  
 Gln Gln Met Leu Ala Asp Ser Ile Asn Arg Met Lys Ala Tyr Ala  
 185 190 195  
 Phe His Gln Lys Lys Glu Ser Met Lys Lys Lys Gln Asp Gly Glu  
 200 205 210  
 Ile Asn Gln Ile Glu Gly Glu Arg Thr Lys Gln Ile Cys Lys Ser  
 215 220 225  
 Trp Lys Glu Asp Ser Glu Trp Gln Ala Ser Leu Arg Lys Ser Lys  
 230 235 240  
 Ala Ala Asp Glu Lys Arg Arg Ser Leu Ala Lys Gln Ala Arg Glu  
 245 250 255  
 Asp Tyr Lys Arg Leu Ser Leu Ala Ala Gln Lys Gly Arg Gly Gly

260	265	270
Glu Arg Leu Gln Ser Pro Leu Arg Val	Pro Gln Lys Pro Glu Arg	
275	280	285
Pro Pro Leu Pro Pro Lys Pro Gln Phe	Leu Asn Ser Gly Ala Tyr	
290	295	300
Pro Gln Lys Pro Leu Arg Asn Gln Gly	Val Val Arg Thr Leu Ser	
305	310	315
Ser Ser Ala Gln Glu Asp Ile Ile Arg	Trp Phe Lys Glu Glu Gln	
320	325	330
Leu Pro Leu Arg Ala Gly Tyr Gln Lys	Thr Ser Asp Thr Ile Ala	
335	340	345
Pro Trp Phe His Gly Ile Leu Thr Leu	Lys Lys Ala Asn Glu Leu	
350	355	360
Leu Leu Ser Thr Gly Met Pro Gly Ser	Phe Leu Ile Arg Val Ser	
365	370	375
Glu Arg Ile Lys Gly Tyr Ala Leu Ser	Tyr Leu Ser Glu Asp Gly	
380	385	390
Cys Lys His Phe Leu Ile Asp Ala Ser	Ala Asp Ala Tyr Ser Phe	
395	400	405
Leu Gly Val Asp Gln Leu Gln His Ala	Thr Leu Ala Asp Leu Val	
410	415	420
Glu Tyr His Lys Glu Glu Pro Ile Thr	Ser Leu Gly Lys Glu Leu	
425	430	435
Leu Leu Tyr Pro Cys Gly Gln Gln Asp	Gln Leu Pro Asp Tyr Leu	
440	445	450
Glu Leu Phe Glu		

&lt;210&gt; 9

&lt;211&gt; 344

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7482809CD1

&lt;400&gt; 9

Met Ala Thr Pro Tyr Val Pro Val Pro Met Pro Ile Gly Asn Ser	
1 5 10 15	
Ala Ser Ser Phe Thr Thr Asn Arg Asn Gln Arg Ser Ser Ser Phe	
20 25 30	
Gly Ser Val Ser Thr Ser Ser Asn Ser Ser Lys Gly Gln Leu Glu	
35 40 45	
Asp Ser Asn Met Gly Asn Phe Lys Gln Thr Ser Val Pro Asp Gln	
50 55 60	
Met Asp Asn Thr Ser Ser Val Cys Ser Ser Pro Leu Ile Arg Thr	
65 70 75	
Lys Phe Thr Gly Thr Ala Ser Ser Ile Glu Tyr Ser Thr Arg Pro	
80 85 90	
Arg Asp Thr Glu Glu Gln Asn Pro Glu Thr Val Asn Trp Glu Asp	
95 100 105	
Arg Pro Ser Thr Pro Thr Ile Leu Gly Tyr Glu Val Met Glu Glu	
110 115 120	
Arg Ala Lys Phe Thr Val Tyr Lys Ile Leu Val Lys Lys Thr Pro	
125 130 135	
Glu Glu Ser Trp Val Val Phe Arg Arg Tyr Thr Asp Phe Ser Arg	
140 145 150	

Leu	Asn	Asp	Lys	Leu	Lys	Glu	Met	Phe	Pro	Gly	Phe	Arg	Leu	Ala	
				155					160					165	
Leu	Pro	Pro	Lys	Arg	Trp	Phe	Lys	Asp	Asn	Tyr	Asn	Ala	Asp	Phe	
				170					175					180	
Leu	Glu	Asp	Arg	Gln	Leu	Gly	Leu	Gln	Ala	Phe	Leu	Gln	Asn	Leu	
				185					190					195	
Val	Ala	His	Lys	Asp	Ile	Ala	Asn	Cys	Leu	Ala	Val	Arg	Glu	Phe	
				200					205					210	
Leu	Cys	Leu	Asp	Asp	Pro	Pro	Gly	Pro	Phe	Asp	Ser	Leu	Glu	Glu	
				215					220					225	
Ser	Arg	Ala	Phe	Cys	Glu	Thr	Leu	Glu	Glu	Thr	Asn	Tyr	Arg	Leu	
				230					235					240	
Gln	Lys	Glu	Leu	Leu	Glu	Lys	Gln	Lys	Glu	Met	Glu	Ser	Leu	Lys	
				245					250					255	
Lys	Leu	Leu	Ser	Glu	Lys	Gln	Leu	His	Ile	Asp	Thr	Leu	Glu	Asn	
				260					265					270	
Arg	Ile	Arg	Thr	Leu	Ser	Leu	Glu	Pro	Glu	Glu	Ser	Leu	Asp	Val	
				275					280					285	
Ser	Glu	Thr	Glu	Gly	Glu	Gln	Ile	Leu	Lys	Val	Glu	Ser	Ser	Ala	
				290					295					300	
Leu	Glu	Val	Asp	Gln	Asp	Val	Leu	Asp	Glu	Glu	Ser	Arg	Ala	Asp	
				305					310					315	
Asn	Lys	Pro	Cys	Leu	Ser	Phe	Ser	Glu	Pro	Glu	Asn	Ala	Val	Ser	
				320					325					330	
Glu	Ile	Glu	Val	Ala	Glu	Val	Ala	Tyr	Asp	Ala	Glu	Glu	Asp		
				335					340						

&lt;210&gt; 10

&lt;211&gt; 1115

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1739178CD1

&lt;400&gt; 10

Met	Ala	Lys	Ile	Val	Val	Val	Thr	Cys	Ser	Asp	Ser	Ser	Phe	Gly	
1				5					10					15	
Asn	Phe	Trp	Leu	Asp	Gln	Trp	Gln	Lys	Arg	Ala	Arg	Glu	Lys	Ser	
				20					25					30	
Leu	Cys	Gln	Cys	Ser	Ala	Lys	Gln	Glu	Ile	Arg	Thr	Gln	Leu	Val	
				35					40					45	
Glu	Gln	Phe	Lys	Cys	Leu	Glu	Gln	Gln	Ser	Glu	Ser	Arg	Leu	Gln	
				50					55					60	
Leu	Leu	Gln	Asp	Leu	Gln	Glu	Phe	Phe	Arg	Arg	Lys	Ala	Glu	Ile	
				65					70					75	
Glu	Leu	Glu	Tyr	Ser	Arg	Ser	Leu	Glu	Lys	Leu	Ala	Glu	Arg	Phe	
				80					85					90	
Ser	Ser	Lys	Ile	Arg	Ser	Ser	Arg	Glu	His	Gln	Phe	Lys	Lys	Asp	
				95					100					105	
Gln	Tyr	Leu	Leu	Ser	Pro	Val	Asn	Cys	Trp	Tyr	Leu	Val	Leu	His	
				110					115					120	
Gln	Thr	Arg	Arg	Glu	Ser	Arg	Asp	His	Ala	Thr	Leu	Asn	Asp	Ile	
				125					130					135	
Phe	Met	Asn	Asn	Val	Ile	Val	Arg	Leu	Ser	Gln	Ile	Ser	Glu	Asp	
				140					145					150	
Val	Ile	Arg	Leu	Phe	Lys	Lys	Ser	Lys	Glu	Ile	Gly	Leu	Gln	Met	

	155		160		165
His Glu Glu Leu	Leu Lys Val Thr Asn Glu	Leu Tyr Thr Val Met			
	170		175		180
Lys Thr Tyr His	Met Tyr His Ala Glu Ser	Ile Ser Ala Glu Ser			
	185		190		195
Lys Leu Lys Glu	Ala Glu Lys Gln Glu Glu	Lys Gln Phe Asn Lys			
	200		205		210
Ser Gly Asp Leu	Ser Met Asn Leu Leu Arg	His Glu Asp Arg Pro			
	215		220		225
Gln Arg Arg Ser	Ser Val Lys Lys Ile Glu	Lys Met Lys Glu Lys			
	230		235		240
Arg Gln Ala Lys	Tyr Ser Glu Asn Lys Leu	Lys Cys Thr Lys Ala			
	245		250		255
Arg Asn Asp Tyr	Leu Leu Asn Leu Ala Ala	Thr Asn Ala Ala Ile			
	260		265		270
Ser Lys Tyr Tyr	Ile His Asp Val Ser Asp	Leu Ile Asp Cys Cys			
	275		280		285
Asp Leu Gly Phe	His Ala Ser Leu Ala Arg	Thr Phe Arg Thr Tyr			
	290		295		300
Leu Ser Ala Glu	Tyr Asn Leu Glu Thr Ser	Arg His Glu Gly Leu			
	305		310		315
Asp Val Ile Glu	Asn Ala Val Asp Asn Leu	Asp Ser Arg Ser Asp			
	320		325		330
Lys His Thr Val	Met Asp Met Cys Asn Gln	Val Phe Cys Pro Pro			
	335		340		345
Leu Lys Phe Glu	Phe Gln Pro His Met Gly	Asp Glu Val Cys Gln			
	350		355		360
Val Ser Ala Gln	Gln Pro Val Gln Thr Glu	Leu Leu Met Arg Tyr			
	365		370		375
His Gln Leu Gln	Ser Arg Leu Ala Thr Leu	Lys Ile Glu Asn Glu			
	380		385		390
Glu Val Arg Lys	Thr Leu Asp Ala Thr Met	Gln Thr Leu Gln Asp			
	395		400		405
Met Leu Thr Val	Glu Asp Phe Asp Val Ser	Asp Ala Phe Gln His			
	410		415		420
Ser Arg Ser Thr	Glu Ser Val Lys Ser Ala	Ala Ser Glu Thr Tyr			
	425		430		435
Met Ser Lys Ile	Asn Ile Ala Lys Arg Arg	Ala Asn Gln Gln Glu			
	440		445		450
Thr Glu Met Phe	Tyr Phe Thr Lys Phe Lys	Glu Tyr Val Asn Gly			
	455		460		465
Ser Asn Leu Ile	Thr Lys Leu Gln Ala Lys	His Asp Leu Leu Lys			
	470		475		480
Gln Thr Leu Gly	Glu Gly Glu Arg Ala Glu	Cys Gly Thr Thr Arg			
	485		490		495
Pro Pro Cys Leu	Pro Pro Lys Pro Gln Lys	Met Arg Arg Pro Arg			
	500		505		510
Pro Leu Ser Val	Tyr Ser His Lys Leu Phe	Asn Gly Ser Met Glu			
	515		520		525
Ala Phe Ile Lys	Asp Ser Gly Gln Ala Ile	Pro Leu Val Val Glu			
	530		535		540
Ser Cys Ile Arg	Tyr Ile Asn Leu Tyr Gly	Leu Gln Gln Gln Gly			
	545		550		555
Ile Phe Arg Val	Pro Gly Ser Gln Val Glu	Val Asn Asp Ile Lys			
	560		565		570
Asn Ser Phe Glu	Arg Gly Glu Asp Pro Leu	Val Asp Asp Gln Asn			
	575		580		585
Glu Arg Asp Ile	Asn Ser Val Ala Gly Val	Leu Lys Leu Tyr Phe			

				590					595				600	
Arg	Gly	Leu	Glu	Asn	Pro	Leu	Phe	Pro	Lys	Glu	Arg	Phe	Gln	Asp
				605					610					615
Leu	Ile	Ser	Thr	Ile	Lys	Leu	Glu	Asn	Pro	Ala	Glu	Arg	Val	His
				620					625					630
Gln	Ile	Gln	Gln	Ile	Leu	Val	Thr	Leu	Pro	Arg	Val	Val	Ile	Val
				635					640					645
Val	Met	Arg	Tyr	Leu	Phe	Ala	Phe	Leu	Asn	His	Leu	Ser	Gln	Tyr
				650					655					660
Ser	Asp	Glu	Asn	Met	Met	Asp	Pro	Tyr	Asn	Leu	Ala	Ile	Cys	Phe
				665					670					675
Gly	Pro	Thr	Leu	Met	His	Ile	Pro	Asp	Gly	Gln	Asp	Pro	Val	Ser
				680					685					690
Cys	Gln	Ala	His	Ile	Asn	Glu	Val	Ile	Lys	Thr	Ile	Ile	Ile	His
				695					700					705
His	Glu	Ala	Ile	Phe	Pro	Ser	Pro	Arg	Glu	Leu	Glu	Gly	Pro	Val
				710					715					720
Tyr	Glu	Lys	Cys	Met	Ala	Gly	Gly	Glu	Glu	Tyr	Cys	Asp	Ser	Pro
				725					730					735
His	Ser	Glu	Pro	Gly	Ala	Ile	Asp	Glu	Val	Asp	His	Asp	Asn	Gly
				740					745					750
Thr	Glu	Pro	His	Thr	Ser	Asp	Glu	Glu	Val	Glu	Gln	Ile	Glu	Ala
				755					760					765
Ile	Ala	Lys	Phe	Asp	Tyr	Met	Gly	Arg	Ser	Pro	Arg	Glu	Leu	Ser
				770					775					780
Phe	Lys	Lys	Gly	Ala	Ser	Leu	Leu	Leu	Tyr	His	Arg	Ala	Ser	Glu
				785					790					795
Asp	Trp	Trp	Glu	Gly	Arg	His	Asn	Gly	Val	Asp	Gly	Leu	Ile	Pro
				800					805					810
His	Gln	Tyr	Ile	Val	Val	Gln	Asp	Met	Asp	Asp	Ala	Phe	Ser	Asp
				815					820					825
Ser	Leu	Ser	Gln	Lys	Ala	Asp	Ser	Glu	Ala	Ser	Ser	Gly	Pro	Leu
				830					835					840
Leu	Asp	Asp	Lys	Ala	Ser	Ser	Lys	Asn	Asp	Leu	Gln	Ser	Pro	Thr
				845					850					855
Glu	His	Ile	Ser	Asp	Tyr	Gly	Phe	Gly	Gly	Val	Met	Gly	Arg	Val
				860					865					870
Arg	Leu	Arg	Ser	Asp	Gly	Ala	Ala	Ile	Pro	Arg	Arg	Arg	Ser	Gly
				875					880					885
Gly	Asp	Thr	His	Ser	Pro	Pro	Arg	Gly	Leu	Gly	Pro	Ser	Ile	Asp
				890					895					900
Thr	Pro	Pro	Arg	Ala	Ala	Ala	Cys	Pro	Ser	Ser	Pro	His	Lys	Ile
				905					910					915
Pro	Leu	Thr	Arg	Gly	Arg	Ile	Glu	Ser	Pro	Glu	Lys	Arg	Arg	Met
				920					925					930
Ala	Thr	Phe	Gly	Ser	Ala	Gly	Ser	Ile	Asn	Tyr	Pro	Asp	Lys	Lys
				935					940					945
Ala	Leu	Ser	Glu	Gly	His	Ser	Met	Arg	Ser	Thr	Cys	Gly	Ser	Thr
				950					955					960
Arg	His	Ser	Ser	Leu	Gly	Asp	His	Lys	Ser	Leu	Glu	Ala	Glu	Ala
				965					970					975
Leu	Ala	Glu	Asp	Ile	Glu	Lys	Thr	Met	Ser	Thr	Ala	Leu	His	Glu
				980					985					990
Leu	Arg	Glu	Leu	Glu	Arg	Gln	Asn	Thr	Val	Lys	Gln	Ala	Pro	Asp
				995					1000					1005
Val	Val	Leu	Asp	Thr	Leu	Glu	Pro	Leu	Lys	Asn	Pro	Pro	Gly	Pro
				1010					1015					1020
Val	Ser	Ser	Glu	Pro	Ala	Ser	Pro	Leu	His	Thr	Ile	Val	Ile	Arg



1025	1030	1035
Asp Pro Asp Ala Ala Met Arg Arg Ser Ser Ser Ser Thr Glu		
1040	1045	1050
Met Met Thr Thr Phe Lys Pro Ala Leu Ser Ala Arg Leu Ala Gly		
1055	1060	1065
Ala Gln Leu Arg Pro Pro Pro Met Arg Pro Val Arg Pro Val Val		
1070	1075	1080
Gln His Arg Ser Ser Ser Ser Ser Ser Ser Gly Val Gly Ser Pro		
1085	1090	1095
Ala Val Thr Pro Thr Glu Lys Met Phe Pro Asn Ser Ser Ala Asp		
1100	1105	1110
Lys Ser Gly Thr Met		
1115		

&lt;210&gt; 11

&lt;211&gt; 839

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473630CD1 .

&lt;400&gt; 11

Met Ala Pro Asn Leu Lys Lys Gly Thr Ser Ser Cys Pro Gly Leu		
1	5	10
15		
Thr Asn Gln Glu Thr His Ser Asp Ser Lys Gly Glu Gly Ala Gly		
20	25	30
Pro Asp Gly Lys Ile Tyr Asp Gly Lys Asp Lys Thr Thr His Leu		
35	40	45
Leu Gly Ala Phe Thr Gly Ala Ser Met Arg Gly Leu Thr Leu Ser		
50	55	60
Ser Thr Ser Asn Gln Leu Trp Leu Glu Phe Asn Ser Asp Thr Glu		
65	70	75
Gly Thr Asp Glu Gly Phe Gln Leu Val Tyr Thr Lys Arg Ile Ile		
80	85	90
Gly Ile Ala Glu Glu Val Thr Val Leu Thr Leu Thr Glu Ser Glu		
95	100	105
Gln Glu Arg Glu His Leu Ser Arg Glu Asp Gln Val Leu Asn Ser		
110	115	120
His Thr Val Lys Ile Leu Ala Phe His Asn Leu Asp Thr Arg Ser		
125	130	135
Val Thr Lys Ala Thr Leu Leu Val Ala Pro Ser Phe Met Asp Ala		
140	145	150
Ile Gln Ala Thr Leu Ser Thr Glu Val Ala Phe Ser Thr Glu Cys		
155	160	165
Gly Gly Arg Phe Lys Gly Glu Ser Ser Gly Arg Ile Leu Ser Pro		
170	175	180
Gly Tyr Pro Phe Pro Tyr Asp Asn Asn Leu Arg Cys Met Trp Met		
185	190	195
Ile Glu Val Asp Pro Gly Asn Ile Val Ser Leu Gln Phe Leu Ala		
200	205	210
Phe Asp Thr Glu Ala Ser His Asp Ile Leu Arg Val Trp Asp Gly		
215	220	225
Pro Pro Glu Asn Asp Met Leu Leu Lys Glu Ile Ser Gly Ser Leu		
230	235	240
Ile Pro Glu Gly Ile His Ser Thr Leu Asn Ile Val Thr Ile Gln		
245	250	255

Phe Asp Thr Asp	Phe Tyr Ile Ser Lys Ser Gly Phe Ala Ile Gln	260	265	270
Phe Ser Ser Ser	Val Ala Thr Ala Cys Arg Asp Pro Gly Val Pro	275	280	285
Met Asn Gly Thr	Arg Asn Gly Asp Gly Arg Glu Pro Gly Asp Thr	290	295	300
Val Val Phe Gln	Cys Asp Pro Gly Tyr Glu Leu Gln Gly Glu Glu	305	310	315
Arg Ile Thr Cys	Ile Gln Val Glu Asn Arg Tyr Phe Trp Gln Pro	320	325	330
Ser Pro Pro Val	Cys Ile Ala Pro Cys Gly Gly Asn Leu Thr Gly	335	340	345
Ser Ser Gly Phe	Ile Leu Ser Pro Asn Phe Pro His Pro Tyr Pro	350	355	360
His Ser Arg Asp	Cys Asp Trp Thr Ile Thr Val Asn Ala Asp Tyr	365	370	375
Val Ile Ser Leu	Ala Phe Ile Ser Phe Ser Ile Glu Pro Asn Tyr	380	385	390
Asp Phe Leu Tyr	Ile Tyr Asp Gly Pro Asp Ser Asn Ser Pro Leu	395	400	405
Ile Gly Ser Phe	Gln Asp Ser Lys Leu Pro Glu Arg Ile Glu Ser	410	415	420
Ser Ser Asn Thr	Met His Leu Ala Phe Arg Ser Asp Gly Ser Val	425	430	435
Ser Tyr Thr Gly	Phe His Leu Glu Tyr Lys Ala Lys Leu Arg Glu	440	445	450
Ser Cys Phe Asp	Pro Gly Asn Ile Met Asn Gly Thr Arg Leu Gly	455	460	465
Met Asp Tyr Lys	Leu Gly Ser Thr Val Thr Tyr Tyr Cys Asp Ala	470	475	480
Gly Tyr Val Leu	Gln Gly Tyr Ser Thr Leu Thr Cys Phe Met Gly	485	490	495
Asp Asp Gly Arg	Pro Gly Trp Asn Arg Ala Leu Pro Ser Cys His	500	505	510
Ala Pro Cys Gly	Ser Arg Ser Thr Gly Ser Glu Gly Thr Val Leu	515	520	525
Ser Pro Asn Tyr	Pro Lys Asn Tyr Ser Val Gly His Asn Cys Val	530	535	540
Tyr Ser Ile Ala	Val Pro Lys Glu Leu Trp Cys Trp Pro Val Val	545	550	555
Phe Phe Gln Thr	Ser Leu His Asp Val Val Glu Val Tyr Asp Gly	560	565	570
Pro Thr Gln Gln	Ser Ser Leu Leu Ser Ser Leu Ser Gly Ser His	575	580	585
Ser Gly Glu Ser	Leu Pro Leu Ser Ser Gly Asn Gln Ile Thr Ile	590	595	600
Arg Phe Thr Ser	Val Gly Pro Ile Thr Ala Lys Gly Phe His Phe	605	610	615
Val Tyr Gln Ala	Val Pro Arg Thr Ser Ser Thr Gln Cys Ser Ser	620	625	630
Val Pro Glu Pro	Arg Phe Gly Arg Arg Ile Gly Asn Glu Phe Ala	635	640	645
Val Gly Ser Ser	Val Leu Phe Asp Cys Asn Pro Gly Tyr Ile Leu	650	655	660
His Gly Ser Ile	Ala Ile Arg Cys Glu Thr Val Pro Asn Ser Leu	665	670	675
Ala Gln Trp Asn	Asp Ser Leu Pro Thr Cys Ile Val Pro Cys Gly	680	685	690

```

Gly Ile Leu Thr Lys Arg Lys Gly Thr Ile Leu Ser Pro Gly Tyr
      695                      700                      705
Pro Glu Pro Tyr Asp Asn Asn Leu Asn Cys Val Trp Lys Ile Thr
      710                      715                      720
Val Pro Glu Gly Ala Gly Ile Gln Val Gln Val Val Ser Phe Ala
      725                      730                      735
Thr Glu His Asn Trp Asp Ser Leu Asp Phe Tyr Asp Gly Gly Asp
      740                      745                      750
Asn Asn Ala Pro Arg Leu Gly Ser Tyr Ser Gly Thr Thr Ile Pro
      755                      760                      765
His Leu Leu Asn Ser Thr Ser Asn Asn Leu Tyr Leu Asn Phe Gln
      770                      775                      780
Ser Asp Ile Ser Val Ser Ala Ala Gly Phe His Leu Glu Tyr Thr
      785                      790                      795
Ala Ile Gly Leu Asp Ser Cys Pro Glu Pro Gln Thr Pro Ser Ser
      800                      805                      810
Gly Ile Lys Ile Gly Asp Arg Tyr Met Val Gly Asp Val Val Ser
      815                      820                      825
Phe Gln Cys Asp Gln Gly Tyr Ser Leu Gln Val Ser Leu Phe
      830                      835

```

&lt;210&gt; 12

&lt;211&gt; 304

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1431520CD1

&lt;400&gt; 12

```

Met Ser Ser Ile Lys His Leu Val Tyr Ala Val Ile Arg Phe Leu
  1      5      10      15
Arg Glu Gln Ser Gln Met Asp Thr Tyr Thr Ser Asp Glu Gln Glu
      20      25      30
Ser Leu Glu Val Ala Ile Gln Cys Leu Glu Thr Val Phe Lys Ile
      35      40      45
Ser Pro Glu Asp Thr His Leu Ala Val Ser Gln Pro Leu Thr Glu
      50      55      60
Met Phe Thr Ser Ser Phe Cys Lys Asn Asp Val Leu Pro Leu Ser
      65      70      75
Asn Ser Val Pro Glu Asp Val Gly Lys Ala Asp Gln Leu Lys Asp
      80      85      90
Glu Gly Asn Asn His Met Lys Glu Glu Asn Tyr Ala Ala Ala Val
      95     100     105
Asp Cys Tyr Thr Gln Ala Ile Glu Leu Asp Pro Asn Asn Ala Val
      110     115     120
Tyr Tyr Cys Asn Arg Ala Ala Ala Gln Ser Lys Leu Gly His Tyr
      125     130     135
Thr Asp Ala Ile Lys Asp Cys Glu Lys Ala Ile Ala Ile Asp Ser
      140     145     150
Lys Tyr Ser Lys Ala Tyr Gly Arg Met Gly Leu Ala Leu Thr Ala
      155     160     165
Leu Asn Lys Phe Glu Glu Ala Val Thr Ser Tyr Gln Lys Ala Leu
      170     175     180
Asp Leu Asp Pro Glu Asn Asp Ser Tyr Lys Ser Asn Leu Lys Ile
      185     190     195
Ala Glu Gln Lys Leu Arg Glu Val Ser Ser Pro Thr Gly Thr Gly

```

	200		205		210
Leu Ser Phe Asp Met	Ala Ser Leu Ile	Asn Asn Pro Ala Phe	Ile		
	215		220		225
Ser Met Ala Ala Ser	Leu Met Gln Asn	Pro Gln Val Gln Gln	Leu		
	230		235		240
Met Ser Gly Met Met	Thr Asn Ala Ile	Gly Gly Pro Ala Ala	Gly		
	245		250		255
Val Gly Gly Leu Thr	Asp Leu Ser Ser	Leu Ile Gln Ala Gly	Gln		
	260		265		270
Gln Phe Ala Gln Gln	Ile Gln Gln Gln	Asn Pro Glu Leu Ile	Glu		
	275		280		285
Gln Leu Arg Asn His	Ile Arg Ser Arg	Ser Phe Ser Ser Ser	Ala		
	290		295		300
Glu Glu His Ser					

&lt;210&gt; 13

&lt;211&gt; 440

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1916304CD1

&lt;400&gt; 13

Met Pro Ile Pro Pro	Pro Pro Pro Pro	Pro Pro Gly Pro Pro	Pro		
1	5	10	15		
Pro Pro Thr Phe His	Gln Ala Asn Thr	Glu Gln Pro Lys Leu	Ser		
	20	25	30		
Arg Asp Glu Gln Arg	Gly Arg Gly Ala	Leu Leu Gln Asp Ile	Cys		
	35	40	45		
Lys Gly Thr Lys Leu	Lys Lys Val Thr	Asn Ile Asn Asp Arg	Ser		
	50	55	60		
Ala Pro Ile Leu Glu	Lys Pro Lys Gly	Ser Ser Gly Gly Tyr	Gly		
	65	70	75		
Ser Gly Gly Ala Ala	Leu Gln Pro Lys	Gly Gly Leu Phe Gln	Gly		
	80	85	90		
Gly Val Leu Lys Leu	Arg Pro Val Gly	Ala Lys Asp Gly Ser	Glu		
	95	100	105		
Asn Leu Ala Gly Lys	Pro Ala Leu Gln	Ile Pro Ser Ser Arg	Ala		
	110	115	120		
Ala Ala Pro Arg Pro	Pro Val Ser Ala	Ala Ser Gly Arg Pro	Gln		
	125	130	135		
Asp Asp Thr Asp Ser	Ser Ser Arg Ala	Ser Leu Pro Glu Leu	Pro Arg		
	140	145	150		
Met Gln Arg Pro Ser	Leu Pro Asp Leu	Ser Arg Pro Asn Thr	Thr		
	155	160	165		
Ser Ser Thr Gly Met	Lys His Ser Ser	Ser Ala Pro Pro Pro	Pro		
	170	175	180		
Pro Pro Gly Arg Arg	Ala Asn Ala Pro	Pro Thr Pro Leu Pro	Met		
	185	190	195		
His Ser Ser Lys Ala	Pro Ala Tyr Asn	Arg Glu Lys Pro Leu	Pro		
	200	205	210		
Pro Thr Pro Gly Gln	Arg Leu His Pro	Gly Arg Glu Gly Pro	Pro		
	215	220	225		
Ala Pro Pro Pro Val	Lys Pro Pro Pro	Ser Pro Val Asn Ile	Arg		
	230	235	240		

```

Thr Gly Pro Ser Gly Gln Ser Leu Ala Pro Pro Pro Pro Pro Tyr
      245      250      255
Arg Gln Pro Pro Gly Val Pro Asn Gly Pro Ser Ser Pro Thr Asn
      260      265      270
Glu Ser Ala Pro Glu Leu Pro Gln Arg His Asn Ser Leu His Arg
      275      280      285
Lys Thr Pro Gly Pro Val Arg Gly Leu Ala Pro Pro Pro Pro Thr
      290      295      300
Ser Ala Ser Pro Ser Leu Leu Ser Asn Arg Pro Pro Pro Pro Ala
      305      310      315
Arg Asp Pro Pro Ser Arg Gly Ala Ala Pro Pro Pro Pro Pro
      320      325      330
Val Ile Arg Asn Gly Ala Arg Asp Ala Pro Pro Pro Pro Pro
      335      340      345
Tyr Arg Met His Gly Ser Glu Pro Pro Ser Arg Gly Lys Pro Pro
      350      355      360
Pro Pro Pro Ser Arg Thr Pro Ala Gly Pro Pro Pro Pro Pro
      365      370      375
Pro Pro Leu Arg Asn Gly His Arg Asp Ser Ile Thr Thr Val Arg
      380      385      390
Ser Phe Leu Asp Asp Phe Glu Ser Lys Tyr Ser Phe His Pro Val
      395      400      405
Glu Asp Phe Pro Ala Pro Glu Glu Tyr Lys His Phe Gln Arg Ile
      410      415      420
Tyr Pro Ser Lys Thr Asn Arg Ala Ala Arg Gly Ala Pro Pro Leu
      425      430      435
Pro Pro Ile Leu Arg
      440

```

&lt;210&gt; 14

&lt;211&gt; 747

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 378504CD1

&lt;400&gt; 14

```

Met Gln Gly Gly Glu Pro Val Ser Thr Met Lys Val Ser Glu Ser
  1      5      10      15
Glu Gly Lys Leu Glu Gly Gln Ala Thr Ala Val Thr Pro Asn Lys
      20      25      30
Asn Ser Ser Cys Gly Gly Ile Ser Ser Ser Ser Ser Arg
      35      40      45
Gly Gly Ser Ala Lys Gly Trp Gln Tyr Ser Asp His Met Glu Asn
      50      55      60
Val Tyr Gly Tyr Leu Met Lys Tyr Thr Asn Leu Val Thr Gly Trp
      65      70      75
Gln Tyr Arg Phe Phe Val Leu Asn Asn Glu Ala Gly Leu Leu Glu
      80      85      90
Tyr Phe Val Asn Glu Gln Ser Arg Asn Gln Lys Pro Arg Gly Thr
      95      100      105
Leu Gln Leu Ala Gly Ala Val Ile Ser Pro Ser Asp Glu Asp Ser
      110      115      120
His Thr Phe Thr Val Asn Ala Ala Ser Gly Glu Gln Tyr Lys Leu
      125      130      135
Arg Ala Thr Asp Ala Lys Glu Arg Gln His Trp Val Ser Arg Leu

```

	140		145		150
Gln Ile Cys Thr	Gln His His Thr Glu	Ala Ile Gly Lys Asn	Asn		
	155		160		165
Pro Pro Leu Lys Ser Arg Ser Phe Ser	Leu Ala Ser Ser Ser	Asn			
	170		175		180
Ser Pro Ile Ser	Gln Arg Arg Pro Ser	Gln Asn Ala Ile Ser	Phe		
	185		190		195
Phe Asn Val Gly	His Ser Lys Leu Gln	Ser Leu Ser Lys Arg	Thr		
	200		205		210
Asn Leu Pro Pro	Asp His Leu Val Glu	Val Arg Glu Met Met	Ser		
	215		220		225
His Ala Glu Gly	Gln Gln Arg Asp Leu	Ile Arg Arg Ile Glu	Cys		
	230		235		240
Leu Pro Thr Ser	Gly His Leu Ser Ser	Leu Asp Gln Asp Leu	Leu		
	245		250		255
Met Leu Lys Ala	Thr Ser Met Ala Thr	Met Asn Cys Leu Asn	Asp		
	260		265		270
Cys Phe His Ile	Leu Gln Leu Gln His	Ala Ser His Gln Lys	Gly		
	275		280		285
Ser Leu Pro Ser	Gly Thr Thr Ile Glu	Trp Leu Glu Pro Lys	Ile		
	290		295		300
Ser Leu Ser Asn	His Tyr Lys Asn Gly	Ala Asp Gln Pro Phe	Ala		
	305		310		315
Thr Asp Gln Ser	Lys Pro Val Ala Val	Pro Glu Glu Gln Pro	Val		
	320		325		330
Ala Glu Ser Gly	Leu Leu Ala Arg Glu	Pro Glu Glu Ile Asn	Ala		
	335		340		345
Asp Asp Glu Ile	Glu Asp Thr Cys Asp	His Lys Glu Asp Asp	Leu		
	350		355		360
Gly Ala Val Glu	Glu Gln Arg Ser Val	Ile Leu His Leu Leu	Ser		
	365		370		375
Gln Leu Lys Leu	Gly Met Asp Leu Thr	Arg Val Val Leu Pro	Thr		
	380		385		390
Phe Ile Leu Glu	Lys Arg Ser Leu Leu	Glu Met Tyr Ala Asp	Phe		
	395		400		405
Met Ser His Pro	Asp Leu Phe Ile Ala	Ile Thr Asn Gly Ala	Thr		
	410		415		420
Ala Glu Asp Arg	Met Ile Arg Phe Phe	Glu Tyr Tyr Leu Thr	Ser		
	425		430		435
Phe His Glu Gly	Arg Lys Gly Ala Ile	Ala Lys Lys Pro Tyr	Asn		
	440		445		450
Pro Ile Ile Gly	Glu Thr Phe His Cys	Ser Trp Lys Met Pro	Lys		
	455		460		465
Ser Glu Val Ala	Ser Ser Val Phe Ser	Ser Ser Ser Thr Gln	Gly		
	470		475		480
Val Thr Asn His	Ala Pro Leu Ser Gly	Glu Ser Leu Thr Gln	Val		
	485		490		495
Gly Ser Asp Cys	Tyr Thr Val Arg Phe	Val Ala Glu Gln Val	Ser		
	500		505		510
His His Pro Pro	Val Ser Gly Phe Tyr	Ala Glu Cys Thr Glu	Arg		
	515		520		525
Lys Met Cys Val	Asn Ala His Val Trp	Thr Lys Ser Lys Phe	Leu		
	530		535		540
Gly Met Ser Ile	Gly Val Thr Met Val	Gly Glu Gly Ile Leu	Ser		
	545		550		555
Leu Leu Glu His	Gly Glu Glu Tyr Thr	Phe Ser Leu Pro Cys	Ala		
	560		565		570
Tyr Ala Arg Ser	Ile Leu Thr Val Pro	Trp Val Glu Leu Gly	Gly		

	575		580		585
Lys Val Ser Val	Asn Cys Ala Lys Thr	Gly Tyr Ser Ala Ser	Ile		
	590		595		600
Thr Phe His Thr	Lys Pro Phe Tyr Gly	Lys Leu His Arg	Val		
	605		610		615
Thr Ala Glu Val	Lys His Asn Ile Thr	Asn Thr Val Val Cys	Arg		
	620		625		630
Val Gln Gly Glu	Trp Asn Ser Val Leu	Glu Phe Thr Tyr Ser	Asn		
	635		640		645
Gly Glu Thr Lys	Tyr Val Asp Leu Thr	Lys Leu Ala Val Thr	Lys		
	650		655		660
Lys Arg Val Arg	Pro Leu Glu Lys Gln	Asp Pro Phe Glu Ser	Arg		
	665		670		675
Arg Leu Trp Lys	Asn Val Thr Asp Ser	Leu Arg Glu Ser Glu	Ile		
	680		685		690
Asp Lys Ala Thr	Glu His Lys His Thr	Leu Glu Glu Arg Gln	Arg		
	695		700		705
Thr Glu Glu Arg	His Arg Thr Glu Thr	Gly Thr Pro Trp Lys	Thr		
	710		715		720
Lys Tyr Phe Ile	Lys Glu Gly Asp Gly	Trp Val Tyr His Lys	Pro		
	725		730		735
Leu Trp Lys Ile	Ile Pro Thr Thr Gln	Pro Ala Glu			
	740		745		

&lt;210&gt; 15

&lt;211&gt; 770

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5275371CD1

&lt;400&gt; 15

Met Pro Leu Leu Glu	Lys Asn Glu Pro Lys	Met Ser Glu Ala Lys	
1	5	10	15
Asn Tyr Leu Ser Ser	Ile Leu Asn His Gly	Arg Leu Ser Pro Gln	
	20	25	30
Tyr Met Cys Glu Ala	Met Leu Ile Leu Gly	Lys Leu His Tyr Val	
	35	40	45
Glu Gly Ser Tyr Arg	Asp Ala Ile Ser Met	Tyr Ala Arg Ala Gly	
	50	55	60
Ile Asp Asp Met Ser	Met Glu Asn Lys Pro	Leu Tyr Gln Met Arg	
	65	70	75
Leu Leu Ser Glu Ala	Phe Val Ile Lys Gly	Leu Ser Leu Glu Arg	
	80	85	90
Leu Pro Asn Ser Ile	Ala Ser Arg Phe Arg	Leu Thr Glu Arg Glu	
	95	100	105
Glu Glu Val Ile Thr	Cys Phe Glu Arg Ala	Ser Trp Ile Ala Gln	
	110	115	120
Val Phe Leu Gln Glu	Leu Glu Lys Thr Thr	Asn Asn Ser Thr Ser	
	125	130	135
Arg His Leu Lys Gly	Cys His Pro Leu Asp	Tyr Glu Leu Thr Tyr	
	140	145	150
Phe Leu Glu Ala Ala	Leu Gln Ser Ala Tyr	Val Lys Asn Leu Lys	
	155	160	165
Lys Gly Asn Ile Val	Lys Gly Met Arg Glu	Leu Arg Glu Val Leu	
	170	175	180

Arg Thr Val Glu Thr Lys Ala Thr Gln Asn Phe Lys Val Met Ala	185	190	195
Ala Lys His Leu Ala Gly Val Leu Leu His Ser Leu Ser Glu Glu	200	205	210
Cys Tyr Trp Ser Pro Leu Ser His Pro Leu Pro Glu Phe Met Gly	215	220	225
Lys Glu Glu Ser Ser Phe Ala Thr Gln Ala Leu Arg Lys Pro His	230	235	240
Leu Tyr Glu Gly Asp Asn Leu Tyr Cys Pro Lys Asp Asn Ile Glu	245	250	255
Glu Ala Leu Leu Leu Leu Leu Ile Ser Glu Ser Met Ala Thr Arg	260	265	270
Asp Val Val Leu Ser Arg Val Pro Glu Gln Glu Glu Asp Arg Thr	275	280	285
Val Ser Leu Gln Asn Ala Ala Ala Ile Tyr Asp Leu Leu Ser Ile	290	295	300
Thr Leu Gly Arg Arg Gly Gln Tyr Val Met Leu Ser Glu Cys Leu	305	310	315
Glu Arg Ala Met Lys Phe Ala Phe Gly Glu Phe His Leu Trp Tyr	320	325	330
Gln Val Ala Leu Ser Met Val Ala Cys Gly Lys Ser Ala Tyr Ala	335	340	345
Val Ser Leu Leu Arg Glu Cys Val Lys Leu Arg Pro Ser Asp Pro	350	355	360
Thr Val Pro Leu Met Ala Ala Lys Val Cys Ile Gly Ser Leu Arg	365	370	375
Trp Leu Glu Glu Ala Glu His Phe Ala Met Met Val Ile Ser Leu	380	385	390
Gly Glu Glu Ala Gly Glu Phe Leu Pro Lys Gly Tyr Leu Ala Leu	395	400	405
Gly Leu Thr Tyr Ser Leu Gln Ala Thr Asp Ala Thr Leu Lys Ser	410	415	420
Lys Gln Asp Glu Leu His Arg Lys Ala Leu Gln Thr Leu Glu Arg	425	430	435
Ala Gln Gln Leu Ala Pro Ser Asp Pro Gln Val Ile Leu Tyr Val	440	445	450
Ser Leu Gln Leu Ala Leu Val Arg Gln Ile Ser Ser Ala Met Glu	455	460	465
Gln Leu Gln Glu Ala Leu Lys Val Arg Lys Asp Asp Ala His Ala	470	475	480
Leu His Leu Leu Ala Leu Leu Phe Ser Ala Gln Lys His His Gln	485	490	495
His Ala Leu Asp Val Val Asn Met Ala Ile Thr Glu His Pro Glu	500	505	510
Asn Phe Asn Leu Met Phe Thr Lys Val Lys Leu Glu Gln Val Leu	515	520	525
Lys Gly Pro Glu Glu Ala Leu Val Thr Cys Arg Gln Val Leu Arg	530	535	540
Leu Trp Gln Thr Leu Tyr Ser Phe Ser Gln Leu Gly Gly Leu Glu	545	550	555
Lys Asp Gly Ser Phe Gly Glu Gly Leu Thr Met Lys Lys Gln Ser	560	565	570
Gly Met His Leu Thr Leu Pro Asp Ala His Asp Ala Asp Ser Gly	575	580	585
Ser Arg Arg Ala Ser Ser Ile Ala Ala Ser Arg Leu Glu Glu Ala	590	595	600
Met Ser Glu Leu Thr Met Pro Ser Ser Val Leu Lys Gln Gly Pro	605	610	615



Met	Gln	Leu	Trp	Thr	Thr	Leu	Glu	Gln	Ile	Trp	Leu	Gln	Ala	Ala	
				620					625					630	
Glu	Leu	Phe	Met	Glu	Gln	Gln	His	Leu	Lys	Glu	Ala	Gly	Phe	Cys	
				635					640					645	
Ile	Gln	Glu	Ala	Ala	Gly	Leu	Phe	Pro	Thr	Ser	His	Ser	Val	Leu	
				650					655					660	
Tyr	Met	Arg	Gly	Arg	Leu	Ala	Glu	Val	Lys	Gly	Asn	Leu	Glu	Glu	
				665					670					675	
Ala	Lys	Gln	Leu	Tyr	Lys	Glu	Ala	Leu	Thr	Val	Asn	Pro	Asp	Gly	
				680					685					690	
Val	Arg	Ile	Met	His	Ser	Leu	Gly	Leu	Met	Leu	Ser	Arg	Leu	Gly	
				695					700					705	
His	Lys	Ser	Leu	Ala	Gln	Lys	Val	Leu	Arg	Asp	Ala	Val	Glu	Arg	
				710					715					720	
Gln	Ser	Thr	Cys	His	Glu	Ala	Trp	Gln	Gly	Leu	Gly	Glu	Val	Leu	
				725					730					735	
Gln	Ala	Gln	Gly	Gln	Asn	Glu	Ala	Ala	Val	Asp	Cys	Phe	Leu	Thr	
				740					745					750	
Ala	Leu	Glu	Leu	Glu	Ala	Ser	Ser	Pro	Val	Leu	Pro	Phe	Ser	Ile	
				755					760					765	
Ile	Pro	Arg	Glu	Leu											
				770											

&lt;210&gt; 16

&lt;211&gt; 199

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 490576CD1

&lt;400&gt; 16

Met	Pro	Glu	Gln	Ser	Asn	Asp	Tyr	Arg	Val	Ala	Val	Phe	Gly	Ala	
1				5					10					15	
Gly	Gly	Val	Gly	Lys	Ser	Ser	Leu	Val	Leu	Arg	Phe	Val	Lys	Gly	
				20					25					30	
Thr	Phe	Arg	Glu	Ser	Tyr	Ile	Pro	Thr	Val	Glu	Asp	Thr	Tyr	Arg	
				35					40					45	
Gln	Val	Ile	Ser	Cys	Asp	Lys	Ser	Ile	Cys	Thr	Leu	Gln	Ile	Thr	
				50					55					60	
Asp	Thr	Thr	Gly	Ser	His	Gln	Phe	Pro	Ala	Met	Gln	Arg	Leu	Ser	
				65					70					75	
Ile	Ser	Lys	Gly	His	Ala	Phe	Ile	Leu	Val	Tyr	Ser	Ile	Thr	Ser	
				80					85					90	
Arg	Gln	Ser	Leu	Glu	Glu	Leu	Lys	Pro	Ile	Tyr	Glu	Gln	Ile	Cys	
				95					100					105	
Glu	Ile	Lys	Gly	Asp	Val	Glu	Ser	Ile	Pro	Ile	Met	Leu	Val	Gly	
				110					115					120	
Asn	Lys	Cys	Asp	Glu	Ser	Pro	Ser	Arg	Glu	Val	Gln	Ser	Ser	Glu	
				125					130					135	
Ala	Glu	Ala	Leu	Ala	Arg	Thr	Trp	Lys	Cys	Ala	Phe	Met	Glu	Thr	
				140					145					150	
Ser	Ala	Lys	Leu	Asn	His	Asn	Val	Lys	Glu	Leu	Phe	Gln	Glu	Leu	
				155					160					165	
Leu	Asn	Leu	Glu	Lys	Arg	Arg	Thr	Val	Ser	Leu	Gln	Ile	Asp	Gly	
				170					175					180	
Lys	Lys	Ser	Lys	Gln	Gln	Lys	Arg	Lys	Glu	Lys	Leu	Lys	Gly	Lys	

Cys Val Ile Met 185 190 195

<210> 17  
 <211> 790  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1417657CD1

<400> 17  
 Met Glu Lys Met Ser Arg Val Thr Thr Ala Leu Gly Gly Ser Val  
 1 5 10 15  
 Leu Thr Gly Arg Thr Met His Cys His Leu Asp Ala Pro Ala Asn  
 20 25 30  
 Ala Ile Ser Val Cys Arg Asp Ala Ala Gln Val Val Val Ala Gly  
 35 40 45  
 Arg Ser Ile Phe Lys Ile Tyr Ala Ile Glu Glu Gln Phe Val  
 50 55 60  
 Glu Lys Leu Asn Leu Arg Val Gly Arg Lys Pro Ser Leu Asn Leu  
 65 70 75  
 Ser Cys Ala Asp Val Val Trp His Gln Met Asp Glu Asn Leu Leu  
 80 85 90  
 Ala Thr Ala Ala Thr Asn Gly Val Val Val Thr Trp Asn Leu Gly  
 95 100 105  
 Arg Pro Ser Arg Asn Lys Gln Asp Gln Leu Phe Thr Glu His Lys  
 110 115 120  
 Arg Thr Val Asn Lys Val Cys Phe His Pro Thr Glu Ala His Val  
 125 130 135  
 Leu Leu Ser Gly Ser Gln Asp Gly Phe Met Lys Cys Phe Asp Leu  
 140 145 150  
 Arg Arg Lys Asp Ser Val Ser Thr Phe Ser Gly Gln Ser Glu Ser  
 155 160 165  
 Val Arg Asp Val Gln Phe Ser Ile Arg Asp Tyr Phe Thr Phe Ala  
 170 175 180  
 Ser Thr Phe Glu Asn Gly Asn Val Gln Leu Trp Asp Ile Arg Arg  
 185 190 195  
 Pro Asp Arg Cys Glu Arg Met Phe Thr Ala His Asn Gly Pro Val  
 200 205 210  
 Phe Cys Cys Asp Trp His Pro Glu Asp Arg Gly Trp Leu Ala Thr  
 215 220 225  
 Gly Gly Arg Asp Lys Met Val Lys Val Trp Asp Met Thr Thr His  
 230 235 240  
 Arg Ala Lys Glu Met His Cys Val Gln Thr Ile Ala Ser Val Ala  
 245 250 255  
 Arg Val Lys Trp Arg Pro Glu Cys Arg His His Leu Ala Thr Cys  
 260 265 270  
 Ser Met Met Val Asp His Asn Ile Tyr Val Trp Asp Val Arg Arg  
 275 280 285  
 Pro Phe Val Pro Ala Ala Met Phe Glu Glu His Arg Asp Val Thr  
 290 295 300  
 Thr Gly Ile Ala Trp Arg His Pro His Asp Pro Ser Phe Leu Leu  
 305 310 315  
 Ser Gly Ser Lys Asp Ser Ser Leu Cys Gln His Leu Phe Arg Asp  
 320 325 330

Ala Ser Gln Pro Val Glu Arg Ala Asn Pro Glu Gly Leu Cys Tyr	335	340	345
Gly Leu Phe Gly Asp Leu Ala Phe Ala Lys Glu Ser Leu Val	350	355	360
Ala Ala Glu Ser Gly Arg Lys Pro Tyr Thr Gly Asp Arg Arg His	365	370	375
Pro Ile Phe Phe Lys Arg Lys Leu Asp Pro Ala Glu Pro Phe Ala	380	385	390
Gly Leu Ala Ser Ser Ala Leu Ser Val Phe Glu Thr Glu Pro Gly	395	400	405
Gly Gly Gly Met Arg Trp Phe Val Asp Thr Ala Glu Arg Tyr Ala	410	415	420
Leu Ala Gly Arg Pro Leu Ala Glu Leu Cys Asp His Asn Ala Lys	425	430	435
Val Ala Arg Glu Leu Gly Arg Asn Gln Val Ala Gln Thr Trp Thr	440	445	450
Met Leu Arg Ile Ile Tyr Cys Ser Pro Gly Leu Val Pro Thr Ala	455	460	465
Asn Leu Asn His Ser Val Gly Lys Gly Gly Ser Cys Gly Leu Pro	470	475	480
Leu Met Asn Ser Phe Asn Leu Lys Asp Met Ala Pro Gly Leu Gly	485	490	495
Ser Glu Thr Arg Leu Asp Arg Ser Lys Gly Asp Ala Arg Ser Asp	500	505	510
Thr Val Leu Leu Asp Ser Ser Ala Thr Leu Ile Thr Asn Glu Asp	515	520	525
Asn Glu Glu Thr Glu Gly Ser Asp Val Pro Ala Asp Tyr Leu Leu	530	535	540
Gly Asp Val Glu Gly Glu Glu Asp Glu Leu Tyr Leu Leu Asp Pro	545	550	555
Glu His Ala His Pro Glu Asp Pro Glu Cys Val Leu Pro Gln Glu	560	565	570
Ala Phe Pro Leu Arg His Glu Ile Val Asp Thr Pro Pro Gly Pro	575	580	585
Glu His Leu Gln Asp Lys Ala Asp Ser Pro His Val Ser Gly Ser	590	595	600
Glu Ala Asp Val Ala Ser Leu Ala Pro Val Asp Ser Ser Phe Ser	605	610	615
Leu Leu Ser Val Ser His Ala Leu Tyr Asp Ser Arg Leu Pro Pro	620	625	630
Asp Phe Phe Gly Val Leu Val Arg Asp Met Leu His Phe Tyr Ala	635	640	645
Glu Gln Gly Asp Val Gln Met Ala Val Ser Val Leu Ile Val Leu	650	655	660
Gly Glu Arg Val Arg Lys Asp Ile Asp Glu Gln Thr Gln Glu His	665	670	675
Trp Tyr Thr Ser Tyr Ile Asp Leu Leu Gln Arg Phe Arg Leu Trp	680	685	690
Asn Val Ser Asn Glu Val Val Lys Leu Ser Thr Ser Arg Ala Val	695	700	705
Ser Cys Leu Asn Gln Ala Ser Thr Thr Leu His Val Asn Cys Ser	710	715	720
His Cys Lys Arg Pro Met Ser Ser Arg Gly Trp Val Cys Asp Arg	725	730	735
Cys His Arg Cys Ala Ser Met Cys Ala Val Cys His His Val Val	740	745	750
Lys Gly Leu Phe Val Trp Cys Gln Gly Cys Ser His Gly Gly His	755	760	765

```

Leu Gln His Ile Met Lys Trp Leu Glu Gly Ser Ser His Cys Pro
              770                      775                      780
Ala Gly Cys Gly His Leu Cys Glu Tyr Ser
              785                      790

```

```
<210> 18
<211> 490
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 1773215CD1
```

<400> 18														
Met	Glu	Glu	Glu	Gly	Val	Lys	Glu	Ala	Gly	Glu	Lys	Pro	Arg	Gly
1				5					10					15
Ala	Gln	Met	Val	Asp	Lys	Ala	Gly	Trp	Ile	Lys	Lys	Ser	Ser	Gly
				20					25					30
Gly	Leu	Leu	Gly	Phe	Trp	Lys	Asp	Arg	Tyr	Leu	Leu	Leu	Cys	Gln
				35					40					45
Ala	Gln	Leu	Leu	Val	Tyr	Glu	Asn	Glu	Asp	Asp	Gln	Lys	Cys	Val
				50					55					60
Glu	Thr	Val	Glu	Leu	Gly	Ser	Tyr	Glu	Lys	Cys	Gln	Asp	Leu	Arg
				65					70					75
Ala	Leu	Leu	Lys	Arg	Lys	His	Arg	Phe	Ile	Leu	Leu	Arg	Ser	Pro
				80					85					90
Gly	Asn	Lys	Val	Ser	Asp	Ile	Lys	Phe	Gln	Ala	Pro	Thr	Gly	Glu
				95					100					105
Glu	Lys	Glu	Ser	Trp	Ile	Lys	Ala	Leu	Asn	Glu	Gly	Ile	Asn	Arg
				110					115					120
Gly	Lys	Asn	Lys	Ala	Phe	Asp	Glu	Val	Lys	Val	Asp	Lys	Ser	Cys
				125					130					135
Ala	Leu	Glu	His	Val	Thr	Arg	Asp	Arg	Val	Arg	Gly	Gly	Gln	Arg
				140					145					150
Arg	Arg	Pro	Pro	Thr	Arg	Val	His	Leu	Lys	Glu	Val	Ala	Ser	Ala
				155					160					165
Ala	Ser	Asp	Gly	Leu	Leu	Arg	Leu	Asp	Leu	Asp	Val	Pro	Asp	Ser
				170					175					180
Gly	Pro	Pro	Val	Phe	Ala	Pro	Ser	Asn	His	Val	Ser	Glu	Ala	Gln
				185					190					195
Pro	Arg	Glu	Thr	Pro	Arg	Pro	Leu	Met	Pro	Pro	Thr	Lys	Pro	Phe
				200					205					210
Leu	Ala	Pro	Glu	Thr	Thr	Ser	Pro	Gly	Asp	Arg	Val	Glu	Thr	Pro
				215					220					225
Val	Gly	Glu	Arg	Ala	Pro	Thr	Pro	Val	Ser	Ala	Ser	Ser	Glu	Val
				230					235					240
Ser	Pro	Glu	Ser	Gln	Glu	Asp	Ser	Glu	Thr	Pro	Ala	Glu	Glu	Asp
				245					250					255
Ser	Gly	Ser	Glu	Gln	Pro	Pro	Asn	Ser	Val	Leu	Pro	Asp	Lys	Leu
				260					265					270
Lys	Val	Ser	Trp	Glu	Asn	Pro	Ser	Pro	Gln	Glu	Ala	Pro	Ala	Ala
				275					280					285
Glu	Ser	Ala	Glu	Pro	Ser	Gln	Ala	Pro	Cys	Ser	Glu	Thr	Ser	Glu
				290					295					300
Ala	Ala	Pro	Arg	Glu	Gly	Gly	Lys	Pro	Pro	Thr	Pro	Pro	Pro	Lys
				305					310					315
Ile	Leu	Ser	Glu	Lys	Leu	Lys	Ala	Ser	Met	Gly	Glu	Met	Gln	Ala

	320		325		330
Ser Gly Pro Pro	Ala Pro Gly Thr Val	Gln Val Ser Val Asn Gly			
	335		340		345
Met Asp Asp Ser	Pro Glu Pro Ala Lys	Pro Ser Gln Ala Glu Gly			
	350		355		360
Thr Pro Gly Thr	Pro Pro Lys Asp Ala	Thr Thr Ser Thr Ala Leu			
	365		370		375
Pro Pro Trp Asp	Leu Pro Pro Gln Phe	His Pro Arg Cys Ser Ser			
	380		385		390
Leu Gly Asp Leu	Leu Gly Glu Gly Pro	Arg His Pro Leu Gln Pro			
	395		400		405
Arg Glu Arg Leu	Tyr Arg Ala Gln Leu	Glu Val Lys Val Ala Ser			
	410		415		420
Glu Gln Thr Glu	Lys Leu Leu Asn Lys	Val Leu Gly Ser Glu Pro			
	425		430		435
Ala Pro Val Ser	Ala Glu Thr Leu Leu	Ser Gln Ala Val Glu Gln			
	440		445		450
Leu Arg Gln Ala	Thr Gln Val Leu Gln	Glu Met Arg Asp Leu Gly			
	455		460		465
Glu Leu Ser Gln	Glu Ala Pro Gly Leu	Arg Glu Lys Arg Lys Glu			
	470		475		480
Leu Val Thr Leu	Tyr Arg Arg Ser Ala	Pro			
	485		490		

&lt;210&gt; 19

&lt;211&gt; 914

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3036986CD1

&lt;400&gt; 19

Met Ala Asn Ile Asn	Leu Lys Glu Ile Thr	Leu Ile Val Gly Val			
1	5	10			15
Val Thr Ala Cys Tyr	Trp Asn Ser Leu Phe	Cys Gly Phe Val Phe			
	20	25			30
Asp Asp Val Ser Ala	Ile Leu Asp Asn Lys	Asp Leu His Pro Ser			
	35	40			45
Thr Pro Leu Lys Thr	Leu Phe Gln Asn Asp	Phe Trp Gly Thr Pro			
	50	55			60
Met Ser Glu Glu Arg	Ser His Lys Ser Tyr	Arg Pro Leu Thr Val			
	65	70			75
Leu Thr Phe Arg Leu	Asn Tyr Leu Leu Ser	Glu Leu Lys Pro Met			
	80	85			90
Ser Tyr His Leu Leu	Asn Met Ile Phe His	Ala Val Val Ser Val			
	95	100			105
Ile Phe Leu Lys Val	Cys Lys Leu Phe Leu	Asp Asn Lys Ser Ser			
	110	115			120
Val Ile Ala Ser Leu	Leu Phe Ala Val His	Pro Ile His Thr Glu			
	125	130			135
Ala Val Thr Gly Val	Val Gly Arg Ala Glu	Leu Leu Ser Ser Ile			
	140	145			150
Phe Phe Leu Ala Ala	Phe Leu Ser Tyr Thr	Arg Ser Lys Gly Pro			
	155	160			165
Asp Asn Ser Ile Ile	Trp Thr Pro Ile Ala	Leu Thr Val Phe Leu			
	170	175			180

Val	Ala	Val	Ala	Thr	Leu	Cys	Lys	Glu	Gln	Gly	Ile	Thr	Val	Val	
				185					190					195	
Gly	Ile	Cys	Cys	Val	Tyr	Glu	Val	Phe	Ile	Ala	Gln	Gly	Tyr	Thr	
				200					205					210	
Leu	Pro	Leu	Leu	Cys	Thr	Thr	Ala	Gly	Gln	Phe	Leu	Arg	Gly	Lys	
				215					220					225	
Gly	Ser	Ile	Pro	Phe	Ser	Met	Leu	Gln	Thr	Leu	Val	Lys	Leu	Ile	
				230					235					240	
Val	Leu	Met	Phe	Ser	Thr	Leu	Leu	Leu	Val	Val	Ile	Arg	Val	Gln	
				245					250					255	
Val	Ile	Gln	Ser	Gln	Leu	Pro	Val	Phe	Thr	Arg	Phe	Asp	Asn	Pro	
				260					265					270	
Ala	Ala	Val	Ser	Pro	Thr	Pro	Thr	Arg	Gln	Leu	Thr	Phe	Asn	Tyr	
				275					280					285	
Leu	Leu	Pro	Val	Asn	Ala	Trp	Leu	Leu	Leu	Asn	Pro	Ser	Glu	Leu	
				290					295					300	
Cys	Cys	Asp	Trp	Thr	Met	Gly	Thr	Ile	Pro	Leu	Ile	Glu	Ser	Leu	
				305					310					315	
Leu	Asp	Ile	Arg	Asn	Leu	Ala	Thr	Phe	Thr	Phe	Phe	Cys	Phe	Leu	
				320					325					330	
Gly	Met	Leu	Gly	Val	Phe	Ser	Ile	Arg	Tyr	Ser	Gly	Asp	Ser	Ser	
				335					340					345	
Lys	Thr	Val	Leu	Met	Ala	Leu	Cys	Leu	Met	Ala	Leu	Pro	Phe	Ile	
				350					355					360	
Pro	Ala	Ser	Asn	Leu	Phe	Phe	Pro	Val	Gly	Phe	Val	Val	Ala	Glu	
				365					370					375	
Arg	Val	Leu	Tyr	Val	Pro	Ser	Met	Gly	Phe	Cys	Ile	Leu	Val	Ala	
				380					385					390	
His	Gly	Trp	Gln	Lys	Ile	Ser	Thr	Lys	Ser	Val	Phe	Lys	Lys	Leu	
				395					400					405	
Ser	Trp	Ile	Cys	Leu	Ser	Met	Val	Ile	Leu	Thr	His	Ser	Leu	Lys	
				410					415					420	
Thr	Phe	His	Arg	Asn	Trp	Asp	Trp	Glu	Ser	Glu	Tyr	Thr	Leu	Phe	
				425					430					435	
Met	Ser	Ala	Leu	Lys	Val	Asn	Lys	Asn	Asn	Ala	Lys	Leu	Trp	Asn	
				440					445					450	
Asn	Val	Gly	His	Ala	Leu	Glu	Asn	Glu	Lys	Asn	Phe	Glu	Arg	Ala	
				455					460					465	
Leu	Lys	Tyr	Phe	Leu	Gln	Ala	Thr	His	Val	Gln	Pro	Asp	Asp	Ile	
				470					475					480	
Gly	Ala	His	Met	Asn	Val	Gly	Arg	Thr	Tyr	Lys	Asn	Leu	Asn	Arg	
				485					490					495	
Thr	Lys	Glu	Ala	Glu	Glu	Ser	Tyr	Met	Met	Ala	Lys	Ser	Leu	Met	
				500					505					510	
Pro	Gln	Ile	Ile	Pro	Gly	Lys	Lys	Tyr	Ala	Ala	Arg	Ile	Ala	Pro	
				515					520					525	
Asn	His	Leu	Asn	Val	Tyr	Ile	Asn	Leu	Ala	Asn	Leu	Ile	Arg	Ala	
				530					535					540	
Asn	Glu	Ser	Arg	Leu	Glu	Glu	Ala	Asp	Gln	Leu	Tyr	Arg	Gln	Ala	
				545					550					555	
Ile	Ser	Met	Arg	Pro	Asp	Phe	Lys	Gln	Ala	Tyr	Ile	Ser	Arg	Gly	
				560					565					570	
Glu	Leu	Leu	Leu	Lys	Met	Asn	Lys	Pro	Leu	Lys	Ala	Lys	Glu	Ala	
				575					580					585	
Tyr	Leu	Lys	Ala	Leu	Glu	Leu	Asp	Arg	Asn	Asn	Ala	Asp	Leu	Trp	
				590					595					600	
Tyr	Asn	Leu	Ala	Ile	Val	His	Ile	Glu	Leu	Lys	Glu	Pro	Asn	Glu	
				605					610					615	

Ala	Leu	Lys	Asn	Phe	Asn	Arg	Ala	Leu	Glu	Leu	Asn	Pro	Lys	His	620	625	630
Lys	Leu	Ala	Leu	Phe	Asn	Ser	Ala	Ile	Val	Met	Gln	Glu	Ser	Gly	635	640	645
Glu	Val	Lys	Leu	Arg	Pro	Glu	Ala	Arg	Lys	Arg	Leu	Leu	Ser	Tyr	650	655	660
Ile	Asn	Glu	Glu	Pro	Leu	Asp	Ala	Asn	Gly	Tyr	Phe	Asn	Leu	Gly	665	670	675
Met	Leu	Ala	Met	Asp	Lys	Lys	Asp	Asn	Glu	Ala	Glu	Ile	Trp		680	685	690
Met	Lys	Lys	Ala	Ile	Lys	Leu	Gln	Ala	Asp	Phe	Arg	Ser	Ala	Leu	695	700	705
Phe	Asn	Leu	Ala	Leu	Leu	Tyr	Ser	Gln	Thr	Ala	Lys	Glu	Leu	Lys	710	715	720
Ala	Leu	Pro	Ile	Leu	Glu	Glu	Leu	Leu	Arg	Tyr	Tyr	Pro	Asp	His	725	730	735
Ile	Lys	Gly	Leu	Ile	Leu	Lys	Gly	Asp	Ile	Leu	Met	Asn	Gln	Lys	740	745	750
Lys	Asp	Ile	Leu	Gly	Ala	Lys	Lys	Cys	Phe	Glu	Arg	Ile	Leu	Glu	755	760	765
Met	Asp	Pro	Ser	Asn	Val	Gln	Gly	Lys	His	Asn	Leu	Cys	Val	Val	770	775	780
Tyr	Phe	Glu	Glu	Lys	Asp	Leu	Leu	Lys	Ala	Glu	Arg	Cys	Leu	Leu	785	790	795
Glu	Thr	Leu	Ala	Leu	Ala	Pro	His	Glu	Glu	Tyr	Ile	Gln	Arg	His	800	805	810
Leu	Asn	Ile	Val	Arg	Asp	Lys	Ile	Ser	Ser	Ser	Ser	Phe	Ile	Glu	815	820	825
Pro	Ile	Phe	Pro	Thr	Ser	Lys	Ile	Ser	Ser	Val	Glu	Gly	Lys	Lys	830	835	840
Ile	Pro	Thr	Glu	Ser	Val	Lys	Glu	Ile	Arg	Gly	Glu	Ser	Arg	Gln	845	850	855
Thr	Gln	Ile	Val	Lys	Thr	Ser	Asp	Asn	Lys	Ser	Gln	Ser	Lys	Ser	860	865	870
Asn	Lys	Gln	Leu	Gly	Lys	Asn	Gly	Asp	Glu	Glu	Thr	Pro	His	Lys	875	880	885
Thr	Thr	Lys	Asp	Ile	Lys	Glu	Ile	Glu	Lys	Lys	Arg	Val	Ala	Ala	890	895	900
Leu	Lys	Arg	Leu	Glu	Glu	Ile	Glu	Arg	Ile	Leu	Asn	Gly	Glu		905	910	

&lt;210&gt; 20

&lt;211&gt; 349

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2041080CD1

&lt;400&gt; 20

Met	Glu	Glu	Glu	Lys	Tyr	Leu	Pro	Glu	Leu	Met	Ala	Glu	Lys	Asp	1	5	10	15
Ser	Leu	Asp	Pro	Ser	Phe	Val	His	Ala	Ser	Arg	Leu	Leu	Ala	Glu	20	25	30	
Glu	Ile	Glu	Lys	Phe	Gln	Gly	Ser	Asp	Gly	Lys	Lys	Glu	Asp	Glu	35	40	45	
Glu	Lys	Lys	Tyr	Leu	Asp	Val	Ile	Ser	Asn	Lys	Asn	Ile	Lys	Leu				

	50		55		60									
Ser	Glu	Arg	Val	Leu	Ile	Pro	Val	Lys	Gln	Tyr	Pro	Lys	Phe	Asn
	65		70		75									
Phe	Val	Gly	Lys	Leu	Leu	Gly	Pro	Arg	Gly	Asn	Ser	Leu	Lys	Arg
	80		85		90									
Leu	Gln	Glu	Glu	Thr	Gly	Ala	Lys	Met	Ser	Ile	Leu	Gly	Lys	Gly
	95		100		105									
Ser	Met	Arg	Asp	Lys	Ala	Lys	Glu	Glu	Glu	Leu	Arg	Lys	Ser	Gly
	110		115		120									
Glu	Ala	Lys	Tyr	Ala	His	Leu	Ser	Asp	Glu	Leu	His	Val	Leu	Ile
	125		130		135									
Glu	Val	Phe	Ala	Pro	Pro	Gly	Glu	Ala	Tyr	Ser	Arg	Met	Ser	His
	140		145		150									
Ala	Leu	Glu	Glu	Ile	Lys	Lys	Phe	Leu	Val	Pro	Asp	Tyr	Asn	Asp
	155		160		165									
Glu	Ile	Arg	Gln	Glu	Gln	Leu	Arg	Glu	Leu	Ser	Tyr	Leu	Asn	Gly
	170		175		180									
Ser	Glu	Asp	Ser	Gly	Arg	Gly	Arg	Gly	Ile	Arg	Gly	Arg	Gly	Ile
	185		190		195									
Arg	Ile	Ala	Pro	Thr	Ala	Pro	Ser	Arg	Gly	Arg	Gly	Gly	Ala	Ile
	200		205		210									
Pro	Pro	Pro	Pro	Pro	Pro	Gly	Arg	Gly	Val	Leu	Thr	Pro	Arg	Gly
	215		220		225									
Ser	Thr	Val	Thr	Arg	Gly	Ala	Leu	Pro	Val	Pro	Pro	Val	Ala	Arg
	230		235		240									
Gly	Val	Pro	Thr	Pro	Arg	Ala	Arg	Gly	Ala	Pro	Thr	Val	Pro	Gly
	245		250		255									
Tyr	Arg	Ala	Pro	Pro	Pro	Pro	Ala	His	Glu	Ala	Tyr	Glu	Glu	Tyr
	260		265		270									
Gly	Tyr	Asp	Asp	Gly	Tyr	Gly	Gly	Glu	Tyr	Asp	Asp	Gln	Thr	Tyr
	275		280		285									
Glu	Thr	Tyr	Asp	Asn	Ser	Tyr	Ala	Thr	Gln	Thr	Gln	Ser	Val	Pro
	290		295		300									
Glu	Tyr	Tyr	Asp	Tyr	Gly	His	Gly	Val	Ser	Glu	Asp	Ala	Tyr	Asp
	305		310		315									
Ser	Tyr	Ala	Pro	Glu	Glu	Trp	Ala	Thr	Thr	Arg	Ser	Ser	Leu	Lys
	320		325		330									
Ala	Pro	Pro	Gln	Arg	Ser	Ala	Arg	Gly	Gly	Tyr	Arg	Glu	His	Pro
	335		340		345									

Tyr Gly Arg Tyr

&lt;210&gt; 21

&lt;211&gt; 2860

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 105283CB1

&lt;400&gt; 21

```

gccagagcgg cggccgggtcc cgcgcgggagc cgggcgcccc tccagcccga gccaggacgc 60
cgccggcccc ggccccggcc cggggcacgc agcgagccag ggatgtgagc ggcccccgc 120
ggcatggcag cctcaggggt gccagagga tgcgacatcc tcatcgtcta cagcccgat 180
gccaggaat ggtgccagta cctgcagacc ctgttcctgt ccagtcggca ggtccgcagc 240
cagaagatac tgactcacag gctgggcccc gaggcctcct tctcggcaga ggacctaagc 300
cttttctca gcaccogctg tgcgtgggtg ctgctgtcog cggagctggt gcagcacttc 360

```



```

cacaagcccg ccttgctgcc cctgctgcag agagctttcc atcctccgca ccgctgggtc 420
aggctgctct gggcgctgag ggacagcgag gagttcctag acttctttcc agattgggcc 480
cattggcagg agctcacctg tgacgatgag ccagagacct acgtggcagc tgtgaaaaaa 540
gccattttccg aagattctgg ctgtgactca gtcactgaca ctgagcctga ggacgagaag 600
gttggtttcct actcgaagca gcagaacctg ccgacgggtga cttcacctgg gaacctgatg 660
gtggtgcagc cggaccgcat tcgctgtggg gcagaaacca ctgtctatgt tattgtgaga 720
tgtaagctgg atgacagggt ggcgacagaa gcagagtttt ctctgagga ttctccctct 780
gtaaggatgg aagccaaggt ggagaatgag tacaccattt cagtgaaggc tcccaacctt 840
tcatctggga acgtttctct gaagatatat tctggagact tagtgggtgtg tgaaaccgtt 900
atcagctatt atactgacat ggaagaaatt gggaatttat tgtccaatgc cgcgaatcct 960
gtggaattca tgtgtcaggc ctttaaaatt gtgcctaca acacagagac ccttgataaa 1020
ctgctaaccg aatccctgaa gaacaatatc cctgcaagcg gactgcacct ctttggaatc 1080
aaccagctgg aagaagaaga tatgatgaca aatcagaggg atgaagagct gccaccctg 1140
ttgcattttg ctgcgaagta tggactgaag aacctcactg ccttggtgct cactgocca 1200
ggagccctgc aggcgtacag cgtggccaac aagcatggcc actaccctaa caccatcgct 1260
gagaaacacg gcttcaggga cctgcggcag ttcacgcagc agtatgtgga aaoggtggac 1320
atgctcaaga gtcacattaa agaggaaactg atgcacgggg aggaggctga tgctgtgtac 1380
gagtcacatg cccacctttc cacagacctg cttatgaaat gctcgtcaa ccccggtgt 1440
gacgaggatc tctatgagtc catggctgac tttgtccag ctgccactga agaccttat 1500
gttgaaatgc ttcaggccag tacatctaac ccaatccctg gagatggttt ctctcgggcc 1560
actaaggact ctatgatccg caagttttta gaaggcaaca gcatgggaat gaccaatctg 1620
gagagagatc agtgccatct tggtcaggaa gaagatgttt atcacacggt ggatgacgat 1680
gaggcctttt ctgtggactt ggccagcagg cccctgttcc cagtgccag accagagacc 1740
actgctcctg gtgctcacca gctgcctgac aacgaaccat acatttttaa agtttttgca 1800
gaaaaaagtc aagagcggcc tgggaatttc tacgtttcct cagagagcat caggaaagg 1860
ccgcccgtca gaccatggag ggacaggccc cagtgcagta tatatgacc ttttgcgga 1920
atgaaaacgc caggccagcg gcagcttacc accctccagg agcaggtgaa gctgggcatt 1980
gtcaacgtgg atgaggctgt gctccacttc aaagagtggc agctcaacca gaagagacga 2040
tcggagtcct ttcgtttcca gcaggaaaat cttaaacggc taagagacag catcacccga 2100
agacagagag agaagcaaaa atcaggaaaag cagacagact tggagatcac ggtcccaatt 2160
cggcactcac agcacctgcc tgcaaaagtg gagtttgagg tctatgagag tggcccag 2220
aaaagtgtca ttccccctag gacggagctg agacgaggag actggaaaac agacagcacc 2280
tccagcacag caagtagcac aagtaaccgc tccagcacc ggagcctcct cagtgtgagc 2340
agcgggatgg aaggggacaa cgaggataat gaagtccctg aggttaccag aagtgcagt 2400
ccaggccccc cacaagtggg tgggacaccc accatgtccc tcgagagacc ccccgagg 2460
cctccgagag ctgcctcaca gaggcctccg accagggaga ccttccatcc tctccacct 2520
gttccacca gaggacgctg attccacctc ctaaaacctg cttacttcag gactttaaga 2580
ctcacagtct tcagcctgtt aatgatgtct tcatgttgag ttttatagca tgactgttga 2640
ccttaagatc cattctcatt gctgataatg ctgcagccct gctggtttgg gcttgccctg 2700
aagattttat taaggcacga agaagtgaag aactaagggc ttcattcacc atcaccaagt 2760
atatcgaacc atatacttgt ttgcaaaaag gatgaagact taatcgaaat acttacctct 2820
aatttgccat atcagaagcc taaaaagaat gatcataaat 2860

```

&lt;210&gt; 22

&lt;211&gt; 3542

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3350821CB1

&lt;400&gt; 22

```

atgtgttgca ctgagggcag tctccggaaa cgcgattcgc agcgggcgcc ggaagcgggtg 60
ttgtgtctgc agctctggca gaggactgtt ccactagaca cgctgaaggg actgggtacg 120
tgttttcctt caggaccaga gctgagagga gctgggatcg cggcggcaat ggaacggggc 180
tcagaaaggc gcacggccag cgcgcttttt gcgggggtcc gggccttggg acttttcagc 240
aacgacattc cacacgtggg gcgggttcagc gcgctcaagc gccggttcta tgtaacaacc 300

```

```

tgcgtgggca agagtttcca cacctatgac gttcagaaac ttagtctggt tgcagtaagt 360
aattctgttc cacaggatat ctgctgtatg gcagctgatg gcagattagt ctttgctgct 420
tatggaaatg ttttctctgc atttgccgt aataaagaga tagtacatac ctttaagggt 480
cataaggcag aaatccattt cttgcaaccc tttggagacc acattatctc tgttgatact 540
gatggcattc ttattatttg gcacatatat tcagaagaag aatacctgca gttgactttt 600
gataaatcag tattttaaatt ttctgcaatt ttgcatccaa gtacctactt gaataaaaata 660
cttctgggca gtgaacaagg aagcctgcag ttgtggaatg taaaatccaa ccagaaatac 720
ccaatcaggc agacatttat ccctgcgggc tatttggttg gtgcacatgg gttgaagact 780
caggcaccag ccgtggatgt tgttgctatt ggtcttatgt caggccaagt tatcattcac 840
aacattaaat ttaatgaaac attaatgaag tttcgtcaag actggggacc cattacttca 900
atctcatttc gcacagatgg tcatccagta atggcagctg gaagcccatg tggccatatt 960
ggactctggg atctagaaga caaaaaatta atcaaccaaa tgagaaatgc acactctaca 1020
gcaattgccg gactgacatt tctccataga gagccacttc ttgtcacaaa tggcgctgac 1080
aatgctctta ggatatggat atttgatggt cctacagggt aaggccgact tttgagattc 1140
agaatgggtc atagtgtccc tcttaccaat atcagatatt atggacagaa tggacagcag 1200
attctaagtg caagtcaaga tggaaactctt cagtcatttt ccacggtaac tgaaaaattc 1260
aataagagct tgggacatgg attaataaat aaaaagagag ttaaacgtaa aggacttcag 1320
aataccattgt cagtggagct tccacccatc acaaagtttg cagcagagga agctcgtgaa 1380
agtgcctggg atggtatcat tgcttgccat caaggtaagc tatcttgctc aacctggaat 1440
tatcagaaat ctacaatagg cgcttacttt ctcaagccaa aagagttgaa gaaagatgac 1500
ataactgcaa cagcagtgga tataactttc ttgtgaaact ttgctgtaat tggcctctca 1560
tcaggaaactg tagatgtata taacatgcag tctggcatac atcgaggaaag ttttggcaag 1620
gatcaagctc acaagggatc tgttagaggt gtcgcagtggt atggattaaa ccagttgaca 1680
gttacaactg gtagtgaagg attactcaaa ttctggaact ttaaaaacaa aattttaatc 1740
cattctgtga gcctcagttc atctccaaat atcatgttgc taatagggga cagtggcatt 1800
ctgggactcg ccttgatga cttctccatt agtgttcttg acatagaagc taggaagatt 1860
gtcagagagt tttctggaca ccaaggccaa ataaatgaca tggcttttag tctgatggt 1920
cgttggttaa taagtgtgc gatggattgc tctattagga cttgggacct tcttctggg 1980
tgcttatag actgctttt gttggactcg gctcctctca atgtttctat gtctcctact 2040
ggagactttc tggcaacttc ccatgtggac caccttggaa tttatctatg gtccaatatt 2100
tccctgtatt cagttgtttc attacggcca ctctctgcag attatgtccc ttcaatagtc 2160
atgcttctctg gtacttgtca aacccaagat gtagaagttt cagaagaaac agtagaacca 2220
agtgatgaat tgatagaata tgattcgcca gaacagttga atgagcaatt ggtgactctt 2280
tcacttcttc ctgaatcacg atggaaaaaac cttcttaacc ttgatgttat taagaaaaag 2340
aataaaccaa aggaaccacc caaagtaccc aaatcagcac catttttcat tccaacaatt 2400
cctggccttg taccagata tgctgcacct gaacaaaata atgatcccca gcagtctaaa 2460
gtggtaaatac ttggagtttt ggctcaaaaa tcagatttct gcttgaaact tgaagaagga 2520
ctggtaaata ataagtatga cactgctctc aaccttctga aagaatcagg cccatcagga 2580
attgaaacag agctgcgaag cttgtctcct gattgtggtg ggtccataga agttatgcag 2640
agcttcttga aaatgattgg gatgatgctg gacagaaagc gtgattttga gttagcccag 2700
gcataccttg cattgtttct aaagttacac cttaaaatgc ttccttcaga gccagtactc 2760
ctagaagaaa taacaaatth gtcatcccag gtggaagaaa actggaccca tttgcaatca 2820
ctcttcaatc aaagcatgtg tattttaaat tatctcaaaa gtgctttgtt gtaaaaaata 2880
atgtgtgact aaacaaagac tttcatatta aatgggttca attgaactca tttcttattt 2940
tccaagtgtc aatgtgaaaa gaaaataaat gctagcacta ctgactagtc agtatatctc 3000
cactttaaat gctaaatact ttcttgaat aaaatcgcac ctcccgcca ggcatgatgg 3060
ataatgcctg taattccagc actttcagag gccaaagcgg gaggattgt tgaagccagg 3120
aatttgagcc cagcctgggc aacatagcaa gcaagatccc atctctacaa aaagttaaaa 3180
aaataaatta cacctgcctt tgttttgaag acttaagaa tctgctcaaa tgatttattt 3240
tctgagtctc ttcacattaa cttttatatg atttttaaaa aattattacc tgcttatatt 3300
tttgagtatc tgtttcatat aaagagatgc tgcagtgtca ctgtgaagta ttgctgaatt 3360
aaattgaaaa tgtcagataa aactgatata tgatgtagtt ataaaaatgc tgttatgaag 3420
aaaaaggaa tgcttttttc ctcggtggcc tttgataaaa atatgaatga ttgactctca 3480
gagtggagga taaactaatc atcttctagc acagccctgt gtgtgcagga gaggaatcag 3540
gg
3542

```

&lt;210&gt; 23

&lt;211&gt; 1014

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5876846CB1

&lt;400&gt; 23

```

ggaccggcg cccgatgaaac cgaccaaact ggagtttagct accggactgt cccgcgagcg 60
aaagttggaa ataataccta aacctacttc ttgacatggt gtcaaaacag agcttttgc 120
gagagtttga tttataaaga aattatgggt cgactgacct tggatctaata tgccagaaac 180
agcaatctta aaccccgaaa agaagaaacc atttcacagt gcctgaagaa aataactcat 240
ataaattttt cagacaaaaa tatagatgca attgaagacc tctctctttg caaaaatctt 300
agtgttttat atttatatga taattgtatt agtcaaata ctaacctgaa ttatgccaca 360
aatctgaccc acttgtagct acaaaacaat tgtatttcat gtatagagaa cctcagggtca 420
ttaaagaaac tggaaaaaact gtatctggga ggcaattaca ttgctgtcat agaaggttta 480
gaaggattag gagaactaag agagcttcat gttgagaatc agaggcttcc ccttggggaa 540
aagcttctgt ttgatccaag aactcttcat tctctggcaa aatccctctg tatattgaat 600
atcagcaata ataattattga tgacattaca gacttagaac tactagagaa tcttaatacag 660
ctcatagccg ttgacaacca acttctgcat gtgaaggatt tggagttttt actgaacaag 720
ttgatgaagc tgtggaaaat tgatctaaat ggaaatcctg tttgtctcaa accaaaatac 780
agggacagac tgatattgggt gtccaaatac ctggaatttc ttgatggaaa agaaattaaa 840
aatatagaaa gacaattcct aatgaattgg aaagcatcca aagatgccaa gaaaatcagc 900
aaaaaaagga gcagtaaaaa tgaggatgca agcaattcac tcataagtaa gcactctgtc 960
actcattagg tgactgctct tctagcaaag gaaatattct aggcaacaaa aaaa 1014

```

&lt;210&gt; 24

&lt;211&gt; 4040

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3560269CB1

&lt;400&gt; 24

```

gcactgcggc acagagccgc ggcttgcttt caggacacgg gtcgctggcg ggtaggtgg 60
ctcgggtggc gcctcacttc gcgcacgcgc cgctaggctg ggggagttga tttgcactgc 120
tccgggtgtc ggctccggcg ggacggggcg ttctgcgtac ccctctccag tgtcaacctg 180
gggctgaatc ttcaacctgt caaagccttc ttaattatgt ctgagaaaag tactaaaata 240
gggtgccattt taaattccca gacaaggaaa aaccaaggat tagaaattaa gtagcttgcc 300
caaggccaca catcagatcc tgattgagtc aatgctttgc attctcagtc ctctgcataa 360
agctgagaga tgcttacagc tgagagtga gcaaaagtaa aaaccaaggt tcgctttgaa 420
gaattgctta agaccacagc tgatctaata cgtgaaaaga aaaaactgaa gaaaaaactt 480
gtcagggtctg aagaaaacat ctacactgac actattagaa gcaatcttca ctatatgaaa 540
gaaactacaa gtgatgatcc cgacactatt agaagcaatc ttccccatat taaagaaact 600
acaagtgatg atgtaagtgc tgctaactc aacaacctga agaagagcac gagagtcact 660
aaaaacaaat tgaggaacac acagttagca actgaaaatc ctaatgggtga tgctagtgt 720
gaggaagaca aacaaggaaa gccaaataaa aaggtgataa agacggtgcc ccagttgact 780
acacaagacc tgaaccggga aactcctgag aataagggtg attctacaca ccagaaaaa 840
catacaagc cacagccagg cgttgatcat cagaaaagtg agaaggcaaa tgagggaaga 900
gaagagactg atttagaaga ggatgaagaa ttgatgcaag catatcagtg ccatgttaact 960
gaagaaatgg caaaggagat taaggagaaa ataagaaaga aactgaaaga acagttgact 1020
tactttccct cagatacttt attccatgat gacaaactaa gcagtgaaaa aaggaaaaag 1080
aaaaaggaag ttccagtctt ctctaaagct gaaacaagta cattgaccat ctctgggtgac 1140
acagttgaag gtgaacaaaa gaaagaatct tcagtttagat cagtttcttc agattctcat 1200
caagatgatg aaataagctc aatggaacaa agcacagaag acagcatgca agatgatata 1260
aaacctaaac caaaaaaac aaaaaagaag actaaagcag ttgcagataa taatgaagat 1320

```

```

gttgatgggtg atgggtgttca tgaataaaca agccgagata gcccgggttta tcccaaatgt 1380
ttgcttgatg atgaccttgt cttgggagtt tacattcacc gaactgatag acttaagtca 1440
gattttatga tttctcacc cttgggtaaaa attcatgtgg ttgatgagca tactgggtcaa 1500
tatgtcaaga aagatgatag tggacggcct gtttcattctt actatgaaaa agagaatgtg 1560
gattatattc ttcctattat gaccagacca tatgatttta aacagttaaa atcaagactt 1620
ccagagtggg aagaacaaat tgtatttaaa gaaaattttc cctatttgct tcgaggctct 1680
gatgagagtc ctaaagtcac cctgttcttt gagattcttg atttcttaag cgtggatgaa 1740
attaagaata attctgaggt tcaaaaccaa gaatgtggct ttcggaaaaat tgcctgggca 1800
tttcttaagc ttcctgggagc caatggaaat gcaaactca actcaaaact tcgcttgagc 1860
ctatattacc cacctactaa gcctcgatcc ccattaagtg ttgttgaggc atttgaatgg 1920
tgggtcaaaat gtccaagaaa tcattaccca tcaactactgt acgtaactgt aagaggactg 1980
aaagttccag actgtataaa gccatcttac cgctctatga tggctcttca ggaggaaaaa 2040
ggtaaacccag tgcattgtga acgtcaccat gagtcaagct cagtagacac agaacctgga 2100
ttagaagagt caaagggaagt aataaagtgg aaacgactcc ctgggcaggc ttgccgtatc 2160
ccaaacaaac acctcttctc actaaatgca ggagaacgag gatgtttttg tcttgatttc 2220
tcccacaatg gaagaatatt agcagcagct tgtgccagcc gggatggata tccaattatt 2280
ttatatgaaa ttccttctgg acgtttcatg agagaattgt gtggccacct caatatcatt 2340
tatgatcttt cctgggtcaaa agatgatcac tacatcctta cttcatcacc tgatggcact 2400
gccaggatat ggaaaaatga aataaacaat acaataactt tcagagtttt acctatcctt 2460
tcttttgttt acacggctaa attccatcca gctgtaagag agctagtagt tacaggatgc 2520
tatgattcca tgatacggat atggaaagtgt gagatgagag aagattctgc catattgggtc 2580
cgacagtttg atgttcacaa aagttttatc aactcacttt gttttgatac tgaagggtcat 2640
catatgtatt caggagattg tacaggggtg attgttgttt ggaataccta tgtcaagatt 2700
aatgatttgg aacattcagt gcaccactgg actataaata aggaaattaa agaaactgag 2760
ttaaggggaa ttccaataag ttatttggag attcatccca atggaaaacg tttgttaatc 2820
cataccaaag acagtacttt gagaattatg gatctccgga tattagtagc aagggaagttt 2880
gtaggagcag caaattatcg ggagaagatt catagtactt tgactccatg tgggactttt 2940
ctgtttgctg gaagtgaagg tggatatagt tacccttcca accagaaac agggaaacaa 3000
gtagccatgt attctgactt gccattcaag tcaccattc gagacatttc ttatcatcca 3060
tttgaaaata tgggtgcatt ctgtgcattt gggcaaaatg agccaattct tctgtatatt 3120
tacgatttcc atgttgccca gcaggaggct gaaatgttca aacgctacaa tgggaacattt 3180
ccattacctg gaatacacca aagtcaagat gccctatgta cctgtccaaa actaccccat 3240
caaggctctt ttcagattga tgaatttgtc cacactgaaa gttcttcaac gaagatgcag 3300
ctagtataaac agaggcttga aactgtcaca gaggtgatac gttcctgtgc tgcaaaagtc 3360
aacaaaaatc tctcatttac ttcaccacca gcagtttctt cacaacagtc taagttaaag 3420
cagtcaaaac tgctgaccgc tcaagagatt ctacatcagt ttggtttcac tcagaccggg 3480
attatcagca tagaaagaaa gccttgtaac catcaggtag atacagcacc aacggtagtg 3540
gctctttatg actacacagc gaatcgatca gatgaactaa ccatccatcg oggagacatt 3600
atccgagtggt ttttcaaaga taatgaagac tgggtggtatg gcagcatagg aaagggacag 3660
gaagggtatt ttocagctaa tcatgtggct agtgaaacac tgtatcaaga actgcctcct 3720
gagataaagg agcgatcccc tcctttaagc cctgaggaaa aaactaaaat agaaaaatct 3780
ccagctcctc aaaaggtaaa ataaaacaaa acagctcaca gagacacctc cttcttcccc 3840
ctgcttctgc ctccatgagt aactacctat atgacgtgct gctgctgaga ccaaggaatg 3900
agttagttaa ggtgtttgga agtcaaatat tggtoctagt taagtaagtg tccttttcaa 3960
gtaagtggta aactttgtta tgtggacccc cttctaactt aggattcata taatttgaac 4020
acaggctaaa gctgctctgg 4040

```

<210> 25

<211> 2006

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 4596874CB1

<400> 25

tccgactggc cgcatacat ctctagagaa gctgacttct ggacagttct caaagaaccg 60

```

cagccgagcc aaacaaggca actcccacca cccccaggc ccctgggagg agtcagagct 120
tgactccaga cccagtggcc atcactggtc gcactgaggg tgaagtggca cggaggagga 180
ggagggggctc cggctgggtct gtggtgagat ggcctaccag gtggtggaga agggcgcgcc 240
cctggggcacg ctggagtcgg agctgcagca gaggcaaagc aggtctggcag ccctggaggc 300
ccgcgtggcg cagctgcgag aggcgcgggc gcagcaggcc cagcaggagg aggagtgggc 360
ggcgcgagaat gcggtgcagc gggcagccta cgaggcgctg cgcgcgcacg tcgggctccg 420
ggagggcgga ctgcgcaggc tccaggaaga ggcgcgcgac ctgctggaga ggctcgtgca 480
gcgcaaggcg cgcgccggc cgcagcgcaa cctgcgcaac gagcgccggg agcggggcaa 540
gcaggcgcgg gtgtcccagg agctgaagaa ggctgccaa cggaccgtga gcatcagcga 600
gggcccggac accctaggcg atgggatgag ggagagaagg gagactctgg ctctggcccc 660
tgagccagag cccctggaga aggaagcttg tgagaagtgg aagaggccct tcagggtctgc 720
ctcagccacc tccctgacgc tgtcccactg tgtggatgtg gtgaaggggc ttctggattt 780
taagaagagg agaggtcact caattggggg agcccctgag cagcgatacc agatcatccc 840
tgtgtgtgtg gctgcccagc ttcctaccgg ggctcaggat gtgctggatg cccacctctc 900
tgaggtcaat gctgttcgtt ttggcccaa cagcagcctc ctggccactg gaggggctga 960
ccgcctgac caccctctga atgttggtgg aagtgcctg gaggccaacc agaccctgga 1020
gggagctggg ggcagcatca ccagtgtgga ctttgacccc tcgggctacc aggttttagc 1080
agcaacttac aaccaggctg cccagctctg gaaggtgggg gaggcacagt ccaaggagac 1140
actgtctgga cacaaggata aggtgacagc tgccaaattc aagctaacga ggcaccaggc 1200
agtactggg agccgcgacc ggacagtga ggagtgggac ctgggccgtg cctattgtctc 1260
caggaccatc aatgtccttt cctactgtaa tgactgggtg tgtggggacc atatcatcat 1320
tagtggccac aatgaccaga agatccgggt ctgggacagc agggggcccc actgcacca 1380
ggtcatccct gtgcagggcc gggtcacctc cctgagcctc agccacgacc aactgcacct 1440
gctcagctgt tcccagaca acacactcaa ggtcatcgac ctgctgttca gcaacatccg 1500
ccaggtgttc agggccgatg gcttcaagtg tggttctgac tggaccaaag ctgtgttcag 1560
cccggacaga agctatgcac tggcaggctc ctgtgatggg gccctttaca tctgggatgt 1620
ggacaccggg aaactggaga gcagactaca gggaccccat tgcgctgccg tcaacgccgt 1680
ggcctgggtc tactccggga gccacatggt gagcgtggac cagggcagga aggttgtgct 1740
ctggcagtag ggccacgacc tgccctgcctg ggctggagct cttgcccga gacctgaagct 1800
tccttcggcg ccatgcaggg gttgggggtg ggactggagc tggccttggg atttaattggg 1860
gaagaaggcc tggcaggacc tggcctgttt gtttaaaaat gaagtatggg ttgggggatt 1920
acgctagttt ttctttgtat ttttatctct atctatctcc tcactttttc tcccaaagta 1980
gaaaaaaatg atatctgaaa aaaaaa
2006

```

&lt;210&gt; 26

&lt;211&gt; 3643

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3594012CB1

&lt;400&gt; 26

```

gaggagaggt acttggtatg attttacagg cgagtcatg caggaaaacc tcagatttgc 60
ttcatcagga gatgatatta aaatatggga tgcttcatct atgacattgg tggataaatt 120
caaccacac acatcaccac atggaatcag ctcaatatgt tggagcagca ataataactt 180
tttagtaaca gcatcttcca gtggcgacaa aatagttgtc tcaagttgca aatgtaaacc 240
tggtccactt ttagagcttg ctgaagggca aaagcagaca tgtgtcaatt taaattctac 300
atctatgtat ttggtaaagc gaggcctaaa taacactgtt aatatttggg atttaaaatc 360
aaaaagagtt catcgatctc ttaaggatca taaagatcaa gtaacttgtg taacatacaa 420
ttggaatgat tgctacattg cttctggatc tcttagtggg gaaattattt tacacagtgt 480
aaccactaat ttatctagta ctcccttttg ccattgtagt aaccagtctg ttcggcactt 540
gaagtactcc ttgtttaaga aatcactact gggcagtggt tcggataatg gaatagtaac 600
tctctgggat gtaaatagtc agagtccata ccataacttt gacagtgtac acaaagctcc 660
agcgtcaggc atctgttttt ctocctgtcaa tgaattgctc tttgtaacca taggcttggg 720
taaaagaatc atcctctatg acacttcaag taagaagcta gtgaaaactt tagtggctga 780
cactcctcta actgcggtag atttcatgcc tgatggagcc actttggcta ttggatcttc 840

```

```

cggggggaaa atatatacaat atgattttaag aatgttgaaa tcaccagtta agaccatcag 900
tgetcacaag acatctgtgc agtgtatagc atttcagtag tccactgttc ttactaagtc 960
aagtttaaat aaaggctgtt caaataagcc cacaacagtg aacaaacgaa gtgttaaatgt 1020
gaatgctgct agtggaggag ttcagaattc cggaattgtc agagaagcac ctgccacgtc 1080
cattgccaca gttctaccac aacctatgac atcagctatg gggaaaaggaa cagttgctgt 1140
tcaagaaaaa gcaggtttgc ctogaagcat aaacacagac actttatcta aggaaacaga 1200
cagtggaaaa aatcaggatt tctccagctt tgatgatact gggaaaagta gtttaggtga 1260
catgttctca cctatcagag atgatgctgt agttaacaag ggaagtgatg agtccatagg 1320
caaaggagat ggctttgact ttctaccgca gttgaactca gtgtttcctc caagaaaaaa 1380
tccagtaact tcaagtactt cagtattgca ttctagtcct cttaatgttt ttatgggatc 1440
tccagggaaa gaggaaaaatg aaaaccgtga tctaacagct gagtctaaga aaatatatat 1500
gggaaaaacag gaatctaaag actccttcaa acagtttagca aagttggtca catctggtgc 1560
tgaaagtgga aatctaaata cctctccatc atctaacca acaagaaatt ctgagaaatt 1620
tgaaaagcca gagaatgaaa ttgaagccca gttgatatgt gaaccccaa tcaatggatc 1680
ctcaactcca aatccaaaga tagcatcttc tgtcactgct ggagttgcca gttcactctc 1740
agaaaaaata gccgacagca ttggaaataa cgggcaaaat gcaccattga cttccattca 1800
aattcgtttt attcagaaca tgatacagga aacgttggat gacttttagag aagcatgcc 1860
tagggacatt gtgaatttgc aagtggagat gattaaacag tttcatatgc aactgaatga 1920
aatgcattct ttgctggaaa gatactcagt gaatgaaggt ttagtggctg aaattgaaag 1980
actacgagaa gaaaacaaaa gattacgggc ccacttttga aatttcagtg aataccttaa 2040
tgttctgtaa tttgggaagt ttctggcaac acagaactac atagaatcag tattgttttc 2100
atggcctcca gggaaaaaat gtttttcaag taagagtaaa egggtgatgg gattttatac 2160
caacaactgt ttcatcttaa aaatatgtat atttttatat taaaaattgt acagtatgtc 2220
atctacccaa taggaaagtc aacaggatct ttattttttg aaagctttag ccatccacta 2280
agtgcccttt ttcataagag aagaaaattg tgcataaaaa ttggttatgt ttgtttttta 2340
gtcatctttt ttaacatata tttttgattg acaaattgcc tttcaaattt ttggggctag 2400
ttgagattta aagagtttga tatgccttct atttttatgg agaaagtaat tttaaaatgg 2460
caattggtgt ttctaagcca ttgactaata aaacataggg ttggctagta attattttgt 2520
taacttgatg aagtcaagta tgactattat ttattgtaca ttgataaga caatttttgg 2580
aattttgaat tgcacaaatt acatgatatc atttgactat atgttactat attgtacttc 2640
tgacaaatct ttattcctgg gtggtatttt taagatatct ttacctataa aaaatgttta 2700
aggttcatag gactcgacaa gagctatctg gtgattttct cattagtaac atgcaacgtt 2760
gtactgcaaa atttcaatca acatgacaac ttataatgag tggagatttc atattaggta 2820
ctaaatatta tagtattatt tctattttct ttttccaaat aagaagcttg gattatttta 2880
ttttgtggtc tttatcatta actttaatte tttctgtact gtgtataata tttttatatt 2940
attggcctta ccataaaatt atttagaaag gttgtcaaaa taagttatac ctctttggca 3000
atagatagat gtatacatct acctactatg atctacaatt ttaggttaag tgaagcttg 3060
gggggctact gacttggtta cttctctgtc tcttggtcca aagattttaa ctatgtacct 3120
ttgtatagct cttctgcccc attttgactt ctgagatgaa agtatttact aaaattaaaa 3180
aaaaaaaac aaaaaacaaa ctttagctc actaaactta tgggtttctg aagtgatgga 3240
aatttttaag gatataattta ataagcataa acttactaat aattacttcc aaaaataaaa 3300
acaggaatat tacttttacc cagtgtggtt tatagcatac atttgtactg aagcatatag 3360
ggatgttaat gtgatctttt cctgacagat tatgaaagca ttatgacttg taacaagttt 3420
ccttgatat cactaacagg tttagaagac ataaatatta gtgtgttttg cctacatgg 3480
gtattttaa ctattaatat tttcctgttg ctttttttaa aaaataaata cacataatgt 3540
atattaaaag aggtgggatg aaataatttt agtaattatg tgtacagatg aaacattttt 3600
gtcatggaat ttaaaagcta agtaagtata aaaaataaaa tgt 3643

```

&lt;210&gt; 27

&lt;211&gt; 2694

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7482435CB1

&lt;400&gt; 27

```

atggcgggaga tccccctgta ctttgtggac ttgcaggatg acttagacga ctatggattt 60
gaagactatg gtacagattg cgacaacatg agagtaacgg ccttcttgga cattccaggc 120
caggataacc tgcctccact cactcgctg gagaaagtatg ctttcagcga gaacaccttc 180
aaccggcgaga ttattgccag agggctgctt gatattcttc gggacttcgg taacaatgaa 240
gaagacttcc tcacggtaat ggagattgta gtcagattgt cagaagatgc agagcccaca 300
gtgcggaactg agctgatgga acagattcct cctattgcca tttttttaca agaaaacaga 360
tcaaattttc cagtgggtgct ctctgaatat ctcataccta ttgtaatgag gtacctcaca 420
gatccaaaca atcagataat ctgcaaaatg gcttcaatgc tgagtaagag taccgttgaa 480
cgctgctac tcccccgatt ctgtgaactg tgtgggtgaca ggaagctctt tcaagttcgg 540
aagggtttgtg ctgcaaaattt tgggtgatatt tgtcatgctg ttggacaaga agccactgaa 600
aaatttttga tcccaaagtt ctttgagctc tgctctgatg ctgtttgggg catgcgga 660
gcctgtgcag aatgcttcac ggctgtgtcg cacagctcct cccctggggg ccgcagaacc 720
cagctcttcc cgctcttcat cagacttgct agcgaccctt gccgatgggt gcaccaggct 780
gccttcagat cccttggccc cttcatttcc accttggcaa acccctocag ggctggcctt 840
tatcttcgag aggatggcgc cctgagcatc tggccactca cccaggattt ggactctggt 900
tttgctctg gctcccctgc tcccagcagt ggtggttaaca tttcccctgc cagtttaatt 960
cgttcagcaa agcctgtgcg gagcgagcca gagctacctg tggaaaggac ctcagcgaaa 1020
accagtgatt gcccacacag cagtagctcc tctgacggcc cagcggaaag cctgttgaa 1080
agctgtgtgt cggtggagc tgagtggacc aggggttccc cagagaccag cgcccgtca 1140
aagctttctg acatgaatga cctcccatc agcagctacc ctggatcaga ttccctgggc 1200
tgccaggga acactgagga tgtgttcagt cattttcttt attgcaagga cttagagctg 1260
cttctgagtg aggtggggcc tcaggaggac gactgcagca gacctggggg tgtgcacaa 1320
agctgtgtgg cccggagtga gattcagaaa gtccttgata gcttgcagga gcatctgatg 1380
aatgatccag atgttcaagc tcaagttcag gtattatccg ctgactgag agctgcacag 1440
ctcgactgcg tgaatgaagc tgagagcaag ccaacagcag gcctaaagga agtgtccatt 1500
tcacatccca gctctgcctc tgacaatcag atcgctctgg cggcctcatc atctcaggat 1560
gagctctttg tggccaggat attacaaagc ccagatccag gtggaccag aaatggaacc 1620
agtgaccatc tggagactga ccagaggcag gatcccacc cacttgaaga gaataaatct 1680
aaattacagg atgtaatacc tcagccgctg ctagatcagt atgtgtccat gactgaccca 1740
gctcgagccc agactgtcga tactgacata gccaaacact gtgcctacag cctcccaggg 1800
gtggcactga ccctgggcag gcaaaattgg cactgcctga aagatacata tgaaacactg 1860
gcttctgatg tacagtggaa ggtacggcga gccctagcct tctccattca cgagctggct 1920
gtgattcttg gggatcagtt aacagcagct gacctgggtc ctatcttcaa tggattttta 1980
aaggatctgg atgaagtgcg aataggagtt cttagacacc tgtatgattt tctaaagttg 2040
cttcatgaag acaaggagg agactatctt tatcagcttc aagaatttgt agtgactgat 2100
aacagcagga attggagggt tcgatatgaa ctagcagaac agctgatact gattctggaa 2160
ctctatagtc ccaatgatgt ttatgattac ctaatgcaca ttgccttaa gttgtgtgca 2220
gatcaagttt ctgaagttcg gtggatctcc ttc aaactag tctgtggcaat tctgcagaag 2280
ttctattcca acagtgaag tgcatgggg ttaaatttca tcaatgagct catcataagg 2340
ttccggcact gttctaagt ggttggaagg caagctttcg ctttcatttg tcaggcagtg 2400
gtgagcaagg agtgtgtccc cgtggaccag ttcattggagc acctgcttcc cagcctcctg 2460
agcctcgcat cagatcctgt gcccaacgtg agggttctgc tagccaaggc cctaaggcag 2520
atgctgttgg aaaaggcgta ttttagaaat gctggttaacc ctcatcttga agtcattgaa 2580
gagaccatct tagcattgca gtcagaccgg gaccaagatg tttcctttt tgcagcccta 2640
gaaccaaaagc ggcggaatat catagacact gctgtactag aaaaacagaa ttaa 2694

```

&lt;210&gt; 28

&lt;211&gt; 2349

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3882333CB1

&lt;400&gt; 28

```

cgccgcttgg gacgcctctg cctttccctc cctcccttcc ccgacggctt ctggcgggcca 60
agtggatgtg gcgggtgatc gagccacct gcccaggggc gccagcact ggttcagtga 120

```

```

acagcatttt ggacaggaca tttggtgcca ggtctgagta gccagtttgc tgaattattg 180
tcccagtcag ccaggattgt gagctgtttg ggaagtttcg tggaaacgcc caagtgccag 240
cacaggtgga gggacacctg gaggccagtt tcaggaactt ttgccacaag tataaaagac 300
ttcagaagtg caaagatgct gaaacagata ctgtcggaga tgtacataga tcttgatcta 360
ctggcagagc tcagcgaaga acagaaacag atcctgttct tcaagatgag agaggaacag 420
atccgacgat ggaaagaaaag agaagcagct atggaaagaa aggagtcctt gccagtgaag 480
cccagaccaa agaaagagaa tggcaaatcg gttcattgga aacttggagc tgataaggaa 540
gtctgggtat ggggtgatggg cgaacaccat ctagataaac cctatgatgt gctctgtaat 600
gaaattattg ctgagagggc cgggctgaaa gcagaacagg aggcagaaga gccagaaaaa 660
actcactctg aagaattcac caatagcttg aaacaaaaat cacagtacca tgatctgcag 720
gctccggata accagcagac taaagacatc tggaaagaaag tggcagaaaa ggaggaactg 780
gagcaaggat ctaggccagc accaaccctg gaagaagaga aaatccgatc actctccagt 840
tcttcaagaa atattcaaca aatgttggca gattcaatca atcgtatgaa ggcatatgca 900
tttcaccaga agaaagaatc tatgaagaaa aaacaagatg gagaaataaa tcaaatagaa 960
ggagagagaa cgaagcagat ttgtaagagc tggaaagaag actcggaatg gcaggcatct 1020
ctgcgaaaat ccaaagcagc tgatgagaag agacgctcct tggctaaaca agcacgagaa 1080
gactacaaga ggttatccct cgcgcccgag aaaggaagag gcggtgagag gctgcaaagc 1140
cccttgctg ttcgcagaa accagaaaga cctcccttc caccgaagcc tcagttccta 1200
aactcaggg catatctca aaaacctctt agaaatcagg gagtggtag gacactgtcc 1260
agctctgccc aagaggacat catccgtgg tttaaagagg agcagctacc acttcgagcg 1320
ggctaccaga aaacctcaga caccatagcc ccctgggttc atggaattct cacactcaag 1380
aaagcaaatg aacttcttct gagcacaggc atgccgggca gttttctcat ccgagtcagt 1440
gaaaggatca aaggctatgc cctgtcctat ctgtcggagg acggtgttaa acatttcctc 1500
atcgatgcct ctgcagacgc ctacagcttc ctgggctgg accagctaca gcatgccacc 1560
ttggcggatt tgggtggaata tcacaaggag gaacccatca cttccctggg gaaggagctc 1620
cttctctatc cctgtggtca gcaggaccag ctgcctgact acctggagct gtttgagtga 1680
cagcctccat cagggtcatc ctacagcctc caagcgggct ttccctgga caaatgccac 1740
tgcaacattt atgtgtgaag ccaaaatcac cctgcagcag agccaatact gatcaactga 1800
aagtaaaagt tccatggagt cctcattgac acctcttttc tgcaaaaata ctggaattca 1860
atgtcaagag aaaatgacct ctgctcaaaa gggagaagag tctcaatttc agcaagtacc 1920
tgtcatgaag ggtatgacct taatgatgta cataaaataa aacaaatgaa gaaatggaaa 1980
acttttagaa attaaggtgt acttgaaaac gaatatotat catatgaccc ctgcaactcc 2040
tctgtatcat ctcaggaggt ttcaggggcc tgttgacatg aagtttcgaa gtttcatgtt 2100
ggctttggaa tggtagcaaa agcctttcct ggctgagatg atgcttaaaa cacacctcac 2160
ttattgtaca tgttggaacc aggacatgag agacatagaa aaacagaagt catgaatgta 2220
aattgaatga gaggtttaac atgcatgaaa atacagatgg acctgcagga aagtgagcaa 2280
acatcgctga gtttgttttc ttgttcggga gaatggggcc ggggctggcc tggcctccct 2340
ggatatact
2349

```

<210> 29

<211> 1213

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7482809CB1

<400> 29

```

aggcgaaaaag ctgcgggttaa ggagagtccg gtttaaccgt caccgggaag cgcgctcggt 60
cgggatcgcc gagtgggctg agatagtga ttcctaagaa gaaaataatg gattgcatat 120
tagttgttct ctaagtggac tcaacagtgt gcaagcttgt tggaaaagcc aaaagaagat 180
ggcaactcct tatgtccag ttcctatgcc cataggaaac tctgcttcca gttttacaac 240
aaacagaaat caaagaagtt cttcttttgg cagtgtctca acaagctcaa attcttctaa 300
gggccagtta gaagactcaa atatgggttaa ttttaaacag acaagtgttc ctgatcaaat 360
ggataatact tcatctgtct gtagcagtc cctcattagg actaaattta cagggtacagc 420
ttcttccatt gagtattcta ctagaccaag agacactgaa gaacaaaatc cggaaacagt 480
gaattgggaa gatagaccat ctacacctac tatactgggt tatgaagtga tggaaagaaag 540

```



```

agctaaat ttt actgtatata aaatactagt aaagaaaacc ccagaagaaa gctgggtagt 600
tttcagaaga tacactgact tctctagggt taatgacaaa ttaaaagaga tgtttccagg 660
ttttcgacta gcaacttcctc caaaacgctg gtttaaagat aattacaatg ctgacttttt 720
agaagacaga caattaggat tacaagcggt tcttcaaaat ttagtagctc acaaggacat 780
tgctaactgc cttgcagtga gagaattttct ttgtttggat gatccaccgg gtccatttga 840
tagccttaga gaaagcaggg cattctgtga aacttttaga gagacaaact accgcttaca 900
gaaagaacta cttgaaaaaac aaaaggagat ggaatcacta aagaaactgc tcagtgaaga 960
gcaacttcat atagacactt tagagaacag aatcagaaca ttgtcttttag aacctgaaga 1020
atcactggat gtgtcagaaa cagaaggtga acagatccta aagggtggagt cctctgcact 1080
tgaggttgat caagatgtcc tggatgaaga atctagagct gataataaac catgcttaag 1140
ttttagtga cctgaaaatg ctgtatcaga gatagaagta gcagaagtgg catatgatgc 1200
tgaagaagac taa 1213

```

&lt;210&gt; 30

&lt;211&gt; 3465

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1739178CB1

&lt;400&gt; 30

```

atggcaaaaa tcgtgggtgt gacatgcagt gattcaagtt ttggaaactt ttggttagat 60
caatggcaaa aacgagcaag agagaaatct ttgtgccaat gctctgccaa gcaagagatc 120
cgcacgcagc tgggtggagca gttcaaatgt ctggagcagc aatcagagtc gcgactgcag 180
ctgcttcaag acctccagga gttttccgc cggaagctg agattgagct cgagtactcc 240
cgcagcctgg agaagctggc tgagcgcttc tcctccaaaa tccgcagctc ccgggagcac 300
cagttcaaga aggaccagta cctcctctcg cctgtgaact gttggtatct ggttctgcat 360
cagacccggc gggagagccg agaccatgcc accctcaatg acatcttcat gaacaatgtc 420
atcgctccgc tctccagat cagtgaggat gtcacagac tcttcaaaaa gagcaaggag 480
attggcctgc agatgcacga ggagctcctg aagggtgacca atgagctcta cacagtcagt 540
aaaacctacc acatgtacca tgcagagagc atcagtgctg aaagcaagct gaaggaggct 600
gagaagcagg aggagaagca gttcaataag tcaggagacc tcagcatgaa cctgctccgg 660
cacgaggacc ggccccagcg ccgcagctct gtgaagaaga ttgagaagat gaaggagaag 720
aggcaggcca agtactctga gaacaagctg aaatgcacaa aggcccgga tgactacttg 780
ctcaatctgg cagccaccaa cgcagctata agcaaatact acatccatga tgtctctgat 840
ctgatcgatt gctgtgattt gggcttccat gccagcctgg ccgcacctt ccggacctat 900
ctctcagctg aatacaacct ggagacotct cgccacgaag ggctggatgt cattgagaat 960
gcagtggaca acctggatc ccgaagtgc aagcacacag tcatggacat gtgcaatcaa 1020
gtcttctgcc ctccactcaa gttcgagttc cagccccaca tgggggatga ggtctgccag 1080
gtcagcgctc agcagccgt ccagacagaa ctgctcatgc gttatcacca gctgcagtcc 1140
agactggcca ccctcaagat agagaatgag gagggttaga aaacctgga tgccaccatg 1200
cagacattac aggacatgct gactgtggag gactttgatg tctccgatgc cttccaacac 1260
agtcgatcga cagagtccgt caagtcggct gcctctgaga cctacatgag caagatcaac 1320
attgccaaaga ggagagccaa ccagcaggaa acagaaatgt tttattttac aaaatttaaa 1380
gagtatgtga atggcagtaa cctcatcacc aagctgcagg ccaagcacga tttactcaag 1440
cagaccctgg gcgaagggga aagagcagaa tgccggcacca ccaggcccc ctgtcttccc 1500
cctaaaccac agaaaatgag gagacctagg cctctctcag tgtatagcca taaactcttt 1560
aacggcagta tggaagcatt cattaaggat tcaggacaag ctataccgct ttagtgcgag 1620
agctgcatcc gttacatcaa tttatatgga ctccagcagc aggggatctt cagagtgcc 1680
ggatctcagg tggaagtcaa tgacatcaaa aattcctttg agagaggtga agacccccctt 1740
gtggacgatc aaaatgaacg agatatcaat tcagtcgctg gtgttttaaa actgtatttc 1800
cgaggactgg aaaacccact ctttcctaag gaaaggtttc aagatttgat atctactatt 1860
aaactggaga acccagccga gagggtgcac cagatccaac aaatcctcgt cacccttccc 1920
cgcgtggcca ttgtggtcat gagatacctc ttcgctttcc tcaaccacct ctccagtat 1980
agcagcaga acatagtgga tccctacaac ttggccatct gcttcggggc taccctcatg 2040
cacatccctg atgggcagga ccctgtgtcc tgccaggcac acatcaatga agtcatcaaa 2100

```

```

accatcatca tccatcatga agccatcttc cccagccccc gggagctaga gggacctgtg 2160
tatgaaaaat gcatggctgg aggggaagaa tattgtgaca gccacacag cgagccaggg 2220
gccatcgcgtg aagttgacca tgacaatggc actgagcctc ataccagcga tgaagaagtg 2280
gagcagatcg aggotattgc caagtttgac tacatggggc ggtccccgcg tgagctatcc 2340
ttcaagaagg gggcctcgct gctcctgtac caccgcgcct cggaggactg gtgggagggc 2400
cggcacaacg gcgtggatgg actcatcccc catcagtaca tagttgtaca ggacatggat 2460
gatgccttct cgcacagcct gagccagaag gctgacagcg aggccagcag tgggccattg 2520
ctggatgaca aggcctcttc caaaaacgac ctccagtccc ccacggagca catctcggat 2580
tacggctttg ggggggtgat gggccgagtg cggttacgat ctgatggagc agccatcccc 2640
agacgccgaa gggggggcga cacacacagc cggccccggg gcctggggcc cagcatagac 2700
acaccacccc gggctgctgc ctgccccagc agccccaca aaatccccct caccgggggg 2760
aggatcgaga gccctgagaa gcggaggatg gcgacgttcg ggagcgtctg cagcatcaac 2820
taccctgaca agaaggcgct ctccgaaggg cactcgatga ggtcgacctg cggttccacc 2880
aggcacagca gcctagggga ccacaagtc ctggaggccg aggccttggc agaagacatc 2940
gagaagacca tgagcacggc tctgcacgag ttgcgggaaac tcgagaggca gaacacggtc 3000
aagcaggcgc cagatgtggt gctggacacc ctggagcccc tgaagaacct gccaggcccc 3060
gtcagctcgg agcccgccag tccccctcac accatcgtca tccgcgacct cgatgccgcc 3120
atgcgcgcga gcagcagctc ctccaccgag atgatgacca ccttcaagcc agccctgtcc 3180
gcccgcctgg ctggcgccca gctccgcccc ccccccattg gggccgtgcg gccggtggtc 3240
cagcacccgt ccagcagcag cagcagctcg ggcgtgggca gcccggccgt gacgccacc 3300
gagaagatgt tccccaacag ctacgcggac aagtcgggca ccatgtgacc tgcaggatgg 3360
gccgcgcgcg cgtggcccg cgtggctcac caccggccag ggtggccttg gtggcttcca 3420
cgtgcttccc agtgaccctg gccttgagg gcagaggctc gtctg 3465

```

&lt;210&gt; 31

&lt;211&gt; 2609

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7473630CB1

&lt;400&gt; 31

```

gtcggcaaca ggctcagccc taccctcaga agcgccctaa aaaacctaca gtgtagaaga 60
gcacaccaga cagtgtgtct gcaagaagga tggctcctaa tctgaaaaaa ggtacaagtt 120
cctgtccagg gctcaccaat caggaactc attctgactc caagggtgaa ggagcagggc 180
cagatggcaa gatttatgat ggaaaagata aaacgactca tctactaggt gcttttactg 240
gtgcatctat ggcgggactg acacttagta gtacttcaaa tcaactctgg ctagaattta 300
attccgatac tgaagggaca gatgaaggct ttcaacttgt gtataccaag agaatcatag 360
gtattgcaga ggaggtcaca gtactcacat tgacagaatc agaacaggag cgagaacatt 420
tgtctaggga ggatcagggt ttgaactctc acactgtgaa gatcctggca ttccacaatt 480
tggatacaag atcagtgacc aaggccactt tgctggtagc accatcattt atggatgcaa 540
tccaggctac actctccacg gaagtagcct totcaactga atgtggaggt cgttttaaa 600
gagaatcatc aggaagaatc ttatctcctg gctatccttt tccatatgac aataacctgc 660
gttgcatgtg gatgattgag gtagatcctg gaaatattgt cagcttgagc tttcttgctt 720
ttgatacgga agcatcacat gatatactcc gagtctggga cggtcacca gaaaatgata 780
tgcttttaaa ggaaattagt ggatctctta ttcctgaagg aattcatagc acctcaata 840
tagtaaccat ccagtttgac acggattttt atattagcaa atctggattt gcaattcagt 900
tttcaagttc tgttgccact gcgtgtcgtg acccaggggt ccccatgaat gggactcgaa 960
atggggatgg aagagaacct ggggacactg ttgtttttca atgtgaccca ggatatgaac 1020
ttcaaggaga ggaaagaata acctgcattc aggtagaaaa tcggtacttc tggcagccca 1080
gcccaccagt ctgtatagca ccctgtggag gcaatttaac aggatcttca ggctttatc 1140
tttcacaaa cttccctcat ccataccgc atagcagaga ctgtgactgg actatcacg 1200
tcaatgcaga ctatgttatc tcttgccgt tcactagttt tagcatagaa ccaactatg 1260
acttctcta tatcatgat ggaccagaca gtaatagcc actgattgga agttttcaag 1320
acagcaagtt accagagaga atagaaagca gctcaaatac aatgcatttg gcttttcgga 1380
gtgatggatc tgttagttac actggatttc atctagaata caaagcaaaa ctgcgagagt 1440

```

```

cctgctttga tccaggcaat ataatgaatg gcaccagact tggaatggat tataaattag 1500
ggtoaacagt cacctattac tgtgatgctg gttatgttct tcaaggttat tcaacactca 1560
cctgtttcat gggagatgat ggaagacctg gatggaatag agccttgcca agttgtcatg 1620
cgccctgtgg aagtcgttca acagggttcag aaggcactgt tctatcacca aactatccaa 1680
aaaattacag tgtgggacat aatttgtgtt attctatagc agttccaaag gagttgtggt 1740
gttggccagt tgtatttttc cagacatcac tccacgatgt tgttgagggt tatgatgggc 1800
caactcagca atcttctctg ttatcttccc tctcaggatc ccattcagga gaatcacttc 1860
cactgagttc aggtaatcag atcacaatc gatttacttc agttggacca ataacagcta 1920
agggatttca ctttgtttac caagctgttc ctagaacaag ttctacacaa tgcagttctg 1980
tgccctgaacc aagattcgga agaagaattg gcaatgaatt tgcagtcggg tcatcggttc 2040
tttttgattg taatccagga tatattctcc atggatccat agcaattagg tgtgaaacag 2100
tgcccaattc tttggcccg tggaatgatt ccttacctac ttgtattgtg ccctgtgggtg 2160
gaattttaac taagcgcaaa gggactattt tgtcacctgg ataccctgag ccttatgaca 2220
acaatctgaa ttgtgtgtgg aagatcacag tgccagaggg agctggcatt caagtgcag 2280
ttgttagctt tgctacagaa cataattggg attctctgga cttttatgat gggggagaca 2340
acaatgctcc aagacttgga agctattcag gaacaacaat accccatctt ttgaatagta 2400
cgtctaataa tctgtatcta aattttcaat cagacatcag tgtttctgct gcaggatttc 2460
atcttgaata cacagcaatt ggtttggatt cctgtcctga accacaaact cctagcagtg 2520
gaattaaaat tggagacaga tatatggttg gagatgtagt atcctttcag tgtgatcaag 2580
gatattctct tcaggtaagt ctattttaa 2609

```

&lt;210&gt; 32

&lt;211&gt; 2580

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1431520CB1

&lt;400&gt; 32

```

gagagagagc acagcctggg gggttctggg gtctacggcc taggggcccgg ggaagtttgc 60
gccgccgcga ccagtgtctg gatcccagc cgggctccag ccccgaggac caggggtcgg 120
gcgggcctgc ctacgggaacc ccgcgggcca gcagcagtcg tctcgctcc tctgtcttg 180
aaaagtgttt aagcttctaa aatgtcatct atcaagcacc tggtttatgc agttattcgt 240
ttcttacggg aacaaagtca gatggacact tacacctcgg atgaacaaga aagtttgaa 300
gttgcaattc agtgcttgga gacagttttt aagatcagcc cagaagatac acacctagca 360
gtttcacagc ctttgacaga aatgtttacc agttccttct gtaagaatga cgtttgcgc 420
ctttcaaac ctgtgctga agatgtggga aaagctgacc aattaaaaga tgaaggcaat 480
aaccacatga aagaagaaaa ttatgctgct gcagtggtt gttacacaca ggcaatagaa 540
ttggatccca ataatgcagt ttactattgc aacagggctg ctgctcagag caaattaggt 600
cactacacag atgogataaa ggattgtgaa aaagcaatag caattgattc aaagtacagc 660
aaggcctatg ggagaatggg gctggccctc actgccttga ataaatttga agaagcagtt 720
acaagttatc aaaaggcatt agatcttgac cctgaaaatg attcctataa gtcaaactctg 780
aaaaatagcag aacagaagtt aagagaggta tccagtccta caggaaactgg actgagcttt 840
gacatggcta gcttgataaa taatccagcc ttcattagta tggcggcaag tttaatgcag 900
aaccctcaag ttcaacagct aatgtcagga atgatgacaa atgccattgg gggacctgct 960
gttgagttg ggggcctaac tgacctgtca agcctcatcc aagcgggaca gcagtttgct 1020
cagcagatac agcaacaaaa tcctgaactt atagagcaac tgagaaatca catccggagc 1080
agatcattca gcagcagcgc tgaagagcat tcotgattta accaggggct caagcccaag 1140
atacaaatgg tttatggcta tgaatgaagt atttggtgta gatagtaccc cctccctcct 1200
tcaaaaaacc aaacaccata tctgatgtgt gaaaataatg gaggaaaacc tcttggtgat 1260
attcttaaag aataaatctg tttagacaat aggtttatca caaacaact caaacataa 1320
catgactcct tttgaaatga tcactagcca tgttacgtgc caatatagca ctgggataag 1380
gttccatttt caaacttcta tcatatttgt atattagact taatagcaaa gcatatttat 1440
tgaactctag tagggctatt tgttggttaag ctcttttagct ccaatttccc ttaagcaag 1500
ttttttgaga agggaaatga cccaaaggaa aggtttccta tgctgggtga gaagaagtgt 1560
acttgccact gagcagtatg tcaggaaaaga aggaatatta tctaggttta gctttgataa 1620

```

```

gtgctattag taatgaatat ataacatggg aaccaatggt atctttaatg ttgcttggtc 1680
tggatgaacag agaatcttaa gagctgttag aaagtagcac ttgatgcaag ggatgttttg 1740
aaaaaagaaa attggtaatg cgaatgtata gaaagtaaag gtaggatgct caggttttct 1800
gcatagttct taactaatct tgtctgcagt ttggtattga taatattagc atggccactt 1860
atgctaaata cacaataaga tacatttaga aatccttaat gtactggtta ggtcagtggt 1920
acaactgttt gacttaatta tcacaatttc cccaatggta accttacctt ggaaactatc 1980
aaatataaat atactttaac taagccagtg acctgggtcct aagttaagga ttttctaaat 2040
ggtatttaat tgtgtgcccc taaacactga ttttttttaa aaagaaaact tagtaattat 2100
ttgatggata acaaaaagca tctatcacta acattgcttg ttcattgtaat ctgtgataat 2160
ttggcttggc atgccagtct tataattaac ataatttaa tttgatgtaa ttttaataata 2220
acaggagact taaaatattt tattataaat tagatgccat attagttagt ggagaaaagg 2280
ggcattaggt attaatgtca tgggtctaac tacattgatg atgagagttt aattccaact 2340
agagtataat ttttgtttat taataatgaa attgggtgctg acatgtaatg acttcacaag 2400
ttttacgagg gaaactgaa aaaggaagta atataaattt tgtgggggtt ttttgagatg 2460
gagtcttgct cttttgccag gctggagtgg tgttgcaatg tcagctcact gcaacctcca 2520
cctcctggat tcaagtgatt ctcgagcctt aagctcccaa gtagctggga ttacgggcgc 2580

```

<210> 33

<211> 2181

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1916304CB1

<400> 33

```

cggcgggcggc gacggcgaga aagagcttgc cggggggcgga gcaggacagg acgaagccgg 60
agtgtaggcg gcagaggatt cgctcccaga gcagctgcgg ccaggtcggg aagaggccgg 120
ggcggtctggg tatatgaatg acctaaaggt acaataaag acggagagag aacagtggca 180
actgggagca gggcaagaat gccaatcct cctccccgc cccccacc tggtcctcct 240
ccacctcca catttcatca ggcaaacaca gagcagcca agctgagtag agatgagcag 300
cggggtcgag gcgcctctct acaggacatt tgcaaggga ccaagctgaa gaaggtgac 360
aacattaatg atcggagtgc tccatcctc gagaagccga aaggaagcag tgggtggctat 420
ggctctggag gagctgccct gcagcccaag ggaggtctct tccaaggagg agtgctgaag 480
cttcgacctg tgggagccaa ggatggttca gagaacctag ctggtaaagg agccctgcaa 540
atccccagtt ctcgagctgc tgcccgaagg cctccagtat ctgccgccag cggggtcct 600
caggatgata cagacagcag cggggcctca ctcccagaac tgccccggat gcagagacct 660
tctttaccgg acctctctcg gcctaatacc accagagta cgggcatgaa gcacagctcc 720
tctgccccct cccaccacc cccagggcgg cgtgccaacg caccoccac acctctgcct 780
atgcacagca gcaagcccc cgctacaac agagagaaac ccttgccacc gacgcctgga 840
caaaggcttc acctggtcg agaggacct cctgctccac cccagtcaa accacctcct 900
tcccctgtga atatcagaac aggaccaagt ggccagtctc tggctcctcc tctcgcct 960
taccgccagc ctctgggggt ccccaatgga cctctagcc ccactaatga gtcagccct 1020
gagctgccac agagacacaa ttctttgcat aggaagacac cagggcctgt cagaggccta 1080
gcacctctc caccacctc ggctcccca tctttactga gtaataggcc acctcccca 1140
gcccagacc ctccagtcg gggagcagct cctccacccc caccacctgt gatccgaaat 1200
ggtgccaggg atgtccccc tccccacca ccataccgaa tgcatgggtc agaacccccg 1260
agccgaggaa agccccacc tccacctca aggacgccag ctgggccacc cctcctcct 1320
ccaccgcccc tgaggaatgg ccacagagat tctatcacca ctgtccggtc tttcttggat 1380
gattttgagt caaagtattc cttccatcca gtagaagact ttctgctcc agaagaatat 1440
aaacactttc agaggatata tcccagcaa acaaaccgag ctgcccgtgg agccccacct 1500
ctgccacca ttctcaggtg aagcctggct tggctccgtt cctcaggaaa aggatggacc 1560
ttctcttct ctccagtggt cccttccatt cccctgaaac ctgcatgaga gtccttaaca 1620
tgtttctcca atgcaatcaa gccctagact ccaaagtcc tcccagctca cctccatcta 1680
tgcatctcat ctctggattt ggtgatcaga ctctatattg acagtaggat ctcaaaccct 1740
gcatccatcc ttctccagc aagcctgct agccacatga ggaacaagtt tccgtgtctt 1800
ctgccttct ctgggggaaa ggtgccttgt tgtgatgaat taactcactg ttagggcagg 1860

```

```

gtggagaatg gtactccttc cttctcctgt ccactgtggg ggaagcttgg caggtatatt 1920
atatttcac attaggagg ctggcatgac caggacttat ggggtggagg ggagcatttt 1980
tagtgaagca agaaaggagt ttgccaagaa gtgatctgtt ttaaagggtca tatttggaga 2040
aagggcaagg aattgggtct gctttatttt tgggggtatt ttgtttttgt tctcacctgc 2100
tgcccccca cccaccacc ccagggataa attggatata aacactaaat actaatcagt 2160
tgaacttaac atttaataaa a                                     2181

```

&lt;210&gt; 34

&lt;211&gt; 4149

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 378504CB1

&lt;400&gt; 34

```

atttgacagt cccgccccag ccgccgtgtg atttagcaac cggatcataa ttcagatcaa 60
atcttaaaaa aaaaaagtca cctacagaac tacagaatgg ttctctcccg ctcaagaagg 120
gcaagtggac ttgggcgtta agatgcaggg ggggtgaacca gtgtccacaa tgaaagtctc 180
ggagagcgaa ggaagctgg agggccaggc cacagcgggtg accccgaaca agaacagcag 240
ctgtggaggt ggaatcagta gcagcagcag cagccgcggt ggcagtgcac aaggctggca 300
gtacagtgat cacatggaaa atgtgtatgg ctatttaatg aagtatacca accttgtcac 360
tgggtggcag tacaggtttt ttgttttaaa caatgaagct gggctgttgg agtactttgt 420
gaatgaacag tctagaaatc agaaacctag aggaaacttg cagcttgcag gagctgtaat 480
atcacccagt gatgaggatt ctacacacct cactgtaaac gctgccagtg gggaacaata 540
taaactcaga gctacagatg caaaagagcg acagcactgg gttagcagac ttcagatag 600
tacacagcat catactgaag ctattggaaa gaataatcct cctctgaagt cacggagctt 660
ctcacttgca tctagtagta attctcctat atcgagagg agaccaagtc aaaaagccat 720
ttcttttttt aatgttggac attccaaact gcaatcactg agcaaaagaa ctaatttacc 780
tccagaccat cttgtggaag tcagagaaat gatgtctcat gctgaaggac aacaaagaga 840
cttaattaga cgaattgaat gccttcctac ttctggccat cttagttcct tggaccagga 900
tctcttaatg ctcaaagcta cttccatggc aactatgaac tgcttaaag actgctttca 960
tattctccag ttacagcatg catcacatca gaagggctca ttgccttcag gaacgacaat 1020
cgagtgggta gaaccaaaga tatctttatc aaaccactat aaaaatggag ctgaccagcc 1080
ctttgcaact gatcagagta agccgggtggc agtccagaa gagcagcctg ttgcagaatc 1140
tggactatta gcgaggagc ctgaagaaat aaatgcagat gatgagatag aggatacatg 1200
tgaccacaaa gaggatgacc tgggagctgt agaagaacaa cgtagtgtca tcctacatct 1260
cttgtcacag cttaagctgg gcatggattt aacaagagtg gtgcttccta catttatcct 1320
agagaagcgt tccttgctgg aaatgtatgc agactttatg tctcatccag acctatttat 1380
agccatcact aatggagcca cagctgagga cagaatgatt cgcttttttg agtactacct 1440
tacctcattt catgaaggcc gtaagggagc cattgctaaa aaaccataca atcctatcat 1500
tggagaaaca tttcactggt cctggaagat gccaaaaagc gaggtagcat ccagtgtttt 1560
tagcagttct tccaccagg gagtcacaaa tcatgctcct ttatcggggg agtctttgac 1620
ccaggtggga tcagactggt acacagtcag atttgttgct gagcagggtt ctcatcatcc 1680
tccagtctca ggattttatg cagaatgtac agagaggaag atgtgtgtaa atgcgcatgt 1740
ctggactaag agcaagttct taggcatgtc aataggcgtg acaatgggtg gagaaggat 1800
ccttagtctg ttggagcatg gagaagagta cacattttct ctaccctgtg catatgctcg 1860
gtcaattttg actgttcctt gggtagaact ggggtggcaaa gtcagtgtca actgtgcaaa 1920
aactggatat tcagccagca tcaactttca taccaagcca ttttatgggt gcaaactgca 1980
tcgggttaca gctgaagtaa agcacaacat caccaacact gtggtatgca gagtgaagg 2040
ggaatggaat agtgttcttg agttcacata tagcaatgga gagacaaagt atgtggactt 2100
gactaaattg gcagtgcaga agaaaagagt gagacctctg gagaagcagg atccatttga 2160
atccaggcga ttgtggaaaa atgtgacaga ctgctgaga gaatctgaaa ttgataaggc 2220
cacagagcat aagcataccc tggagaagc tcagaggact gaagaaaggc atcgactatga 2280
aacaggcaca ccttggaaaa ccaaatattt tattaaagag ggagatggct gggtttatca 2340
taaaccactt tggaaaaataa ttccaacaac acaaccagca gagtgcacac tactatctaa 2400
aactcgacca aatgaggttc ttctctgttt accctaaatc ctcccagaat ggagtcattg 2460

```

```

cactgagtga cctgcttctt gattgcgcag actgaaacta gctaaacctg aatgtacctt 2520
ctagggcacc ataatactgc agcaagacca aagtggtaaa gaaacacagt ggacctttta 2580
ccaacctgtt catgtgatgt gagcaatacc atcttaaaac ttgttacctg aatcagtaga 2640
tgaatctttt atccagttct tgctcctaaa gttaagtttg aatccctat ttttgcacag 2700
gggcagcaga tacacacaac aatgagaact cagtgaactt gatttctttg tagtgaaaag 2760
tgaagtctcc gtttcagagt ttgtgtcttt ctttctgtcc ataactgaag tattcactac 2820
tcttgtaaac caaccaagag gaggagaaag atgaccacaga agtggattca gccattgtgc 2880
ctgaaatcag tgtttaaaaa aaaaaatcaa ccaggttgtg gtaacaaggc attctatttc 2940
ttcaaaaaga ctgtatgcct gtgtctgagg aacttaccta ttatccacct ctgttggaaac 3000
tctcttttaa aaagtacatt tatagattga tcagaattat aacctgggag aattttttct 3060
tctgagcatt ttaatactat tgaaaacaac attgacttga aaaaattcag aacatttttc 3120
agtacctagt tttattaaat attacacttg agagacactt tttaaaaatg tgtaaatgtc 3180
aatatgatga gatttttagcc tttctccaga actaaggcat taaagaaaat agcaaataat 3240
aaaaataaaa actgttactt ttttcttctt ttcttttcac ctttaggtta atatccagta 3300
ttatgtgtta tccctttgga taagtatgct ttattttacc tctgttaaaa attaaaaata 3360
atgattctat tcatatttgt cagtaattca aaacttatat gtgtaactga acgogcatgt 3420
aaggtatggt tttatttatt tttttttttt ttgaggaaat ttaaatgcta aagaaacaac 3480
gaaatgaaaa ggtatcagga aaaaaagatc aggaagttgt attcaggtag aaatcttttt 3540
ttaaataagt attttggtga ggttgaagaa ttgctggcaa ttaaaagaat agagctaatt 3600
atggctttca tcatctatc atgtatttat tgagcaccta cttattatgg tgcacaacac 3660
ttgttactgc aagctacctt aatttcccaa gagtgggtgc ttactctgtt ttttctgata 3720
tggtcttcca atcagtgtgt gtaacatacc tgttgtttat cagccattgt aggtggctgt 3780
gtctgttgca tcatcataag aagtttaagc tttgtgctct gataaattgt gttctgttaa 3840
agaggttagt aggatgaaaa cagcaaaaaca ataatttttt caacaaattg taaattataa 3900
gaaaaagagt tggtttgtgt acaacaattt taatgattcc ctgttcatt tttgctgtga 3960
aatgcaactga aaaaaatcct caaaatgagt tatagttcct gtgttgggaa aattgacaaa 4020
taataaaact agagaacaac catataacaa aaaacaaaaa acaacaaaac aaaaaaaaca 4080
acacaacaca acaaaaaaca aaaaaacaaa caaccacaaa caaaaaaaa cagagccccc 4140
catagacca                                     4149

```

<210> 35

<211> 3080

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 5275371CB1

<400> 35

```

gctagaaaatg gggaaaacca attatgggtta catagaaatt aggagcatgt aaaagtacat 60
tattaaacca cttcatgtat ttttctcttt caatcatttg gggagaattc ttcttggttt 120
aagaacttat tatccacagc acagcatatg tatgactttg tcccttattt taaattctga 180
aattctggga atttgtcccc gagtcctggg aactctgtct tcccaaccat taaactacag 240
cagaaaataa atccaagggg caatttgtat tgtgaccata gtagaatgcc tgtgtctttt 300
ctttaaaggc tgttgaaatt tatgaaaaat gtgcctttat tttctcagac ctgagccatt 360
catgtactgg gtatgggtgg ttgggtgggt ccaggagtgt gcaactccct ccattatccg 420
ccttgttcgg ggtttgctgc tctgaccctt actttgcttt tcagatgact ttgggaaatt 480
gctgctggct gaggccctcc tggagcagtg tttgaaggag aacctgcca aaataaaaga 540
ctccatgcct ttgctggaga agaattgagc gaagatgagc gaagccaaaa attatctaag 600
cagtatcctt aaccatggga ggctctcgcc acagtacatg tgtgaggcca tgctgatcct 660
gggcaaactg cattacgtgg agggctcata ccgagatgcc atcagcatgt acgcacgggc 720
cgggattgat gacatgtcca tggagaacaa gccctgtat cagatgcggc tgctgtcgga 780
ggcttttgtc atcaaaggcc tctctctgga acgcctaccc aactccatcg cctcccgctt 840
ccgcctgaca gagagggagg aggaagtgat cacctgtttt gagagggcct cctggatcgc 900
tcaggtgttc ctgcaggaat tggagaagac cacaataaac agcacgtcga ggcatctgaa 960
aggctgtcac ccgcttgact atgagctcac ctacttcctg gaagctgccc tccagagcgc 1020
ctatgtgaaa aacctgaaga aggggaacat cgtgaagggc atgagagagc tccgggaagt 1080

```

```

gctgcggaact gtggagacca aagcaactca gaacttcaaa gtgatggcgg ccaagcacct 1140
ggcggggggtc ctgctgcact cctgagtgga ggagtgctac tggagcccc tgtccaccc 1200
tctgcctgag ttcattggga aggaggagag ttctttcgcc actcaggccc tgcggaaacc 1260
tcacctctat gaaggagaca acctctactg ccccaaggac aacatcgagg aagccctcct 1320
gctcctcctc atcagcgaat ccatggcaac tcgagatgtg gtgctgagcc ggggtgccga 1380
gcaggaggag gaccggacag tgagcttgca gaatgccga gccatctatg acctcctgag 1440
catcacgttg ggcagaagg gacagtacgt catgctctcg gagtgcctgg agcgagccat 1500
gaagtgtgog tttggagaat ttcacctttg gtaccagggt gccctctcca tgggtggctg 1560
tgggaagtca gcctacgctg tgtccctgct gggggagtgt gtgaagtgc ggccctcgga 1620
ccccaccgtg cccctgatgg ccgcgaagg ctgcatcggt tcccttcgtt ggctagagga 1680
agcagagcac tttgccatga tgggtatcag cctcggagag gaagccgggg agttcctccc 1740
caagggtctac ctggctcttg gtctcaccta tagcctgcag gccaccgacg ccaccctgaa 1800
gtccaagcaa gatgaattgc accggaaggc actgcagacg ctggagaggg ctcagcagct 1860
ggcgcccagt gacccccagg tcatcctcta tgtctcgctg cagctggccc tcgtccgaca 1920
gatctccagt gccatggagc agctgcagga ggccctgaag gtacgcaagg atgatgccc 1980
cgccctccac ctgctggcac tgccttctc tgcgcagaag caccaccagc atgccctgga 2040
tggtgtcaac atggccatca ccgagcacc tgagaacttc aacctgatgt tcaccaaggt 2100
gaagctggag caggtgctga aaggcccgga ggaagccctc gtgacctgca gacaagtgt 2160
gaggtctggg cagaccctgt acagcttctc ccagctggga ggcctagaaa aggatggcag 2220
cttcggtgag ggcctcacca tgaagaagca gagtggcatg cacctgactt tgccctgatg 2280
ccatgatgca gactctggct cccggcgggc ttgctccatc gccgcctccc ggctggagga 2340
ggccatgtca gagctgacta tgccctcttc ggtcctgaag caggggccca tgcagctgtg 2400
gaccacgtg gaacagatct ggctgcaggc tgcctgagctg ttcattggagc agcagcacct 2460
caaggaagca ggtttctgca tccaggaggc ggcgggcctc tccccactt ctcactcagt 2520
actctatatg cggggccggc tggctgaggt gaagggcaac ctggaggagg ccaagcagct 2580
gtacaaggag gcgctcacgg tgaaccaga tggcgtgcgc atcatgcata gcctgggtct 2640
gatgctgagt cggctgggccc acaagajctt ggcccagaag gtgcttcgtg atgcccgtga 2700
gaggcagagt acgtgccacg aggcgtggca ggcctgggc gagtgctgc aggccaggg 2760
ccagaacgag gctgccgttg actgcttctt caccgcctt gagctggagg ccagcagccc 2820
tgtactgccc ttctccatca tccccagaga gctctgacga cgctgcagcc gcagggaggg 2880
aggggctggc cagagggaga ggcagcaggg aacgtgggtc aggggtgggc aacagtggca 2940
tcaggtgcgg ggcctcaggg aaatacatct ttagtgaacg cctctgcagc tgcagccctc 3000
gttctcttg ctggggccaa aggccttctt ggatttctt gttgtcaatt gggaacaggc 3060
tgctgagcac cgactgccag                                     3080

```

&lt;210&gt; 36

&lt;211&gt; 4167

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 490576CB1

&lt;400&gt; 36

```

ggcgggcgga gggagtgccg tgcgcctgcc tccgggagga gccgcatcca cacaccctgc 60
gctgccctgt cctgcgcgag tggagctctg aagaagctct gagcggagtt gtgttcttcc 120
ccagggtcgt cctggctgag agttggagct ctccagcaac atgcctgagc agagtaacga 180
ttaccgggtg gccgtgtttg gggctggcgg tgttggcaag agctccctgg tgttgaggtt 240
tgtgaaaggc acattccggg agagctacat cccgacggtg gaagacacct accggcaagt 300
gatcagctgt gacaagagca tatgcacatt gcagatcacc gacacgacgg ggagccaacca 360
gttcccggcc atgcagcggc tgtccatctc caaagggcac gccttcatcc tgggtgtactc 420
cattaccagc cgacagtctt tggaggagct caagcccatc tacgaacaaa tctgcgagat 480
caaaggggac gtggagagca tccccatcat gctgggtggg aacaagtgtg atgagagccc 540
cagccgcgag gtgcagagca gcgaggcgga ggccttggcc cgcacatgga agtgtgcctt 600
catggagacc tcagccaagg tcaaccataa cgtgaaggag cttttccagg agctgctcaa 660
cctggagaag cgcaggaccg tgagtctcca gatcgacggg aaaaagagca agcagcagaa 720
aaggaaagag aagctcaaa gcaagtgcgt gatcatgtga aggccttcc tgcgggagga 780

```

```

gcagctgtgt gtccccggca cctcactccc ccaaaatgac acccacggtc gtcagggtag 840
catgtataat gccacgtgt taaacattgc atttaatcga gatgcgtcct attgtcctta 900
agagggcggt tccaccacc aacagtaagc caccactct ggagtcacag aatctgccag 960
gcggttcaag tgaaaaccaa cacactcagc atccctggga actgagaggt gccagcaatt 1020
gctgaagggt gcgatgaaca cccgaagggt ggagggagga ctggtacca caaagcaaca 1080
tgtaccgaga ggactaaatg tcatctacgt gcatgtgaga gcgtgttaac ctagagttac 1140
ctgcaccaac cccagacaga agccaatcac atctttgggg gaggggaggg gcaggaagag 1200
gtgagaagat cagatgggtc aaagtggacc acacttggtc cattttacac ttttttaaag 1260
gggattaaaa aacacagcct ctccccaaa ggggtgctgt tcttaattcc cactggcct 1320
gttagggacc ttgctaccct gaggggatgt gttcacctta cctagacctt gttaggaagt 1380
atcattttta gctattagag tatttatctt catgtgcagg gataagtgca ctaacagtgt 1440
gctgctctgt oggaagttct tcagttttta agtgaggata tcgtgacagt attaaaacat 1500
cgcaataatg ttctgtgtgt ttatacatcg agggtttttag aaatgtgatt ttcttctttt 1560
gacctgtgag gagtataact tctttcagcc ctacagatttt aaatacaagc aaataaactc 1620
actattttta gacgtttttt tcctccaagg tgggttttctt ctctaaata actcgatctg 1680
taccagctg ggtagcagcc agcaaaggcc atcagacaac cagaagcaca tccatttttg 1740
tagtgtcaca aacatgtata tgccacactt tgcaccttaa tgaataactt tgaacagaa 1800
gttattcact gtgtttttga tgatctatct gtattggaaa tatgttctctg gaaaatgcat 1860
ttaataata gtaaatcttc ttgcatgttc cattatacgt gtcttctaag agctgttcaa 1920
tacagtattc actctagaaa caattatctt tttctcttaa tgattttgtg tgcatcttta 1980
atctttcaag ccaaattaca gctatttcag gtttctctgt ttagcttggg gataggatgg 2040
tggctggaga caggcaggct tctctgccct ggggaagagcc cactcagctt aattgctctg 2100
ccatcgtaga gcctgggttg acttggtctt ctgaaaactc ccactgatag tgctgttag 2160
atctcctgtt tgtttcagtt ggcagaacat ttactggccc caactgtggc atcatcctct 2220
cagcagctct cctgtcaccc gcctggcagg cagaaggagc tgcagtccca cgtgggcctg 2280
cctggggggg tgggggctgc atggctgttg ggtggcagtg tcagcacagg gagggcttaa 2340
gttggggatg ttgaccagg ccacctcctg caactgctgt ttctcctgtc cctcctatgc 2400
agggcttgca gcagcagcag tgtggccatc tccatcccc aaagcacact tgctctctca 2460
atatgtccta gttttcttca gccttttctg gttcagttcc cttgtcctga tctcactctc 2520
tctggtctcc caataactca cccttgggat cgttttagag cgtgggaggt gcctttgaga 2580
actgcttgac tccatgatct cctagaacaa aaccgccctg actttacagg gggaacactc 2640
atgctgagct gagaaagcag agaagtggcg tgggagccag ctgggggtga agagcatttg 2700
ggccagtcct gtggccccct tcagattcct caagcaggat tgttctgttc taaaaagctg 2760
ttgcacagca ttcgcaatga gatctttagt tggcggattt tctggaacat ttgtttttca 2820
acttgtcccg acattttttt ttctgtttct attctgagag agagatgatc aagttttaat 2880
ttgggtatag gttaaattga agaagaaaca gaacttcatg gccaaagtag acctatagat 2940
tttgattggg ttctttgtta acagtagaat gogatctttg ccactgactg tagtattaat 3000
aaggttttaa tgtgagatat tcctgcaaac catcccattt ctactgattg taagtcagaa 3060
ttctttttat ccttttcaaa tcagtttcta catgtttaag tgttcagggc ttcactcaga 3120
tgagaagttt gtaattactg aaagtctgat ttcattcagg acacattttt ttccttcata 3180
ttttttctgt gaatttatag gctaggaagg ctattgaagc ctcaattatg ggtcttcatt 3240
ttgagatcgt ttcttatgag ctgaactgag gatataaat gttatctcaa aatcgtcttt 3300
taggagatcc ccaattgact cagagtttga ggagttagta tcacagaatt agattttttt 3360
aaagcatttg tacgtttcca ttcccaata tgtagctgtg gttcttgaaa acacatccta 3420
cattgcatat gggcatagca gtttttgacc caggcagaat aagttaatat ttaattaaat 3480
attgctttga agatggcgct ctgggcatga gcatggggct ccactgactt ccttctatcc 3540
ccatgagccc ctctccatc cagcgacaag ccactgggcat gcatacaatg cagcaagacc 3600
aacacaagag caatattgaa ttgttcattc tatctaaaat tacatgtata taaaatatat 3660
aatttatctt cctgcatttt tgaagtataa agtcataaat tgtacatatc tgtaagctag 3720
tatatttggt tcaactgttg taatatttta gaaatgctca ttctttgtag aacaaaaatg 3780
tattaaatat tttaaaaatt gctctgtgat acttaatttt tttcccaaaa atttgtaatg 3840
tgttgcttct acataagttc tctggaaata tctacaacta ataggacaca tgtaaatcct 3900
tgaagacaca tcctggaatt cataccccac aaggacagtg tgtatacaaa gtatttgcag 3960
agcatgactt ttatatgtgt gggatatcaa tgtgtatatt tatattttaa gtgtatttat 4020
tgttacaagt ctattctcta ttatatttta tttactctgc ggttataaaa atcacccttg 4080
catacaagtt tctagttgcc agtgatgttc tggaaataat gggagatatt acaataaagc 4140
tacagttatg acaaaaaaaa aaaaaaaa 4167

```



<210> 37  
 <211> 3591  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1417657CB1

<400> 37

```

gggccccctcc tggcctctgc caacgcagct ccccgaccgc gaactttcgt ggtgtcgggc 60
gagagtttgt gcccaacttc tgagagggat gggtaagaag aggaccactt ccggcgaggg 120
acgggaaaga cagcggcttc cggcgcgggc gttccggaca accgtgcgct tttagtaaaa 180
gattgggggtt cgcgcggggg agaagggctg ccccgggccc tctggttctc gtcccgcagc 240
gtccgctccc cgcgcgcaact gcgcgctcc caggaacctt gtactccggg gtcgcgggct 300
tctctcctgc ctccgggtccc gccagacacc tcgagctcct taagtagctc ggctcctgac 360
gtccctctgg gcccttcccc cgtctatcgc ctgagtcctc gggccccctc agccctctgt 420
tccctccccct cttttgttcc tccctagagc cccgcgcgcc tcagggtga cagtgtggac 480
ggcgggagtc tccctcgtcc cctgctggga ttgactgacc gagcgttttag tgactgccc 540
gatctggctg atgggggtac cgagagggtg cctggggccg gaatgtccag ctagagtctt 600
ccgtggaagt cagacatgaa actgacaggc ctaagggaag ctaggaagtc cctcaccgc 660
tcagccaggg tgatgggctg gactgacaga ctccagtga tttgagcttg cctgtcaggc 720
tgattggctg atagacagcc ctggattggc tctaactaac tgaccagccc gggaccaagc 780
agttctgggg tcccaacctg ggtggaaggt ctgaactgat gaccaccca ggctgaccag 840
gccagccac ctactgacc tctgacccc tgacctcatc acctgtgcag ccatggagaa 900
gatgtccggt gtgaccacag ccctgggtgg cagcgtgctg acaggccgca ccatgcaact 960
ccacctggat gctcccgcca atgccatcag tgtgtgccc gacgcagccc aggtggtcgt 1020
ggcaggccgt agcatcttca agatctatgc catcgaggag gaacagtctg tggaaaagct 1080
gaacctgcgt gtggggcgca agccttcgct taacctgagc tgtgctgacg tggctcggca 1140
ccagatggat gagaacctgc tggccacagc agccaccaat ggcgtgggtg tcacgtggaa 1200
cctggggccg ccatcccgca acaagcagga ccagctgttc acagaacaca agcgcacggt 1260
aaacaaagtc tgcttccacc ccaccgaagc ccacgtgctg ctcagtggct cccaggatgg 1320
cttcatgaag tgctttgacc tccgcagaaa ggactctgtc agcaccttct cgggccagtc 1380
ggagagcgtg cgggacgtgc agttcagtat cggggactac ttcaccttcg cctccacctt 1440
tgagaacggc aatgtgcagc tctgggacat cgggcgtccc gaccggtgcg agaggatgtt 1500
cacagcccag aacggacccg tcttctgctg cgaactggac cccgaggaca ggggctgggt 1560
ggccactgga gggcgcgaca agatggtgaa ggtctgggac atgaccacgc accgtgcca 1620
ggagatgcac tgtgtgcaga ccatcgctc ggtggcccg gtgaagtggc ggccagagt 1680
ccgccaccac ctggccacgt gctccatgat ggtggaccac aacatctatg tttgggacgt 1740
gcgcggggcc tctgtgccag ctgccatgtt tgaggaacac cgagacgtca ccacgggaat 1800
tgcctggcgc cccccccacg accctcctt cctgctgtct ggctccaagg acagctcgt 1860
gtgccagcac ctgttccgcg acgccagcca gcccgctcag cgcgccaacc ctgagggcct 1920
ctgctacggc ctcttcgggg acctggcctt cgcgcgcaag gagagcctcg tggctgccga 1980
gtcggggcgc aagccctaca ctggcgaccg gcgccacccc atcttcttta agcgcaagct 2040
ggacctgcc gagcccttcg caggcctcgc ctccagtgc ctcagtgtct ttgagacgga 2100
gccaggtggc ggcggcatgc gctggtttgt ggacacagct gacggttatg cgctggctgg 2160
ccggccactg ggcgagctct gtgaccacaa cgcaaagggt gctcgagagc ttggccgcaa 2220
ccaggtggcg caaacgtgga ccatgctgcg gatcatctac tgcagccctg gcctagtgc 2280
cactgcaaac ctcaaccaca gtgtgggcaa ggggtggctc tgtggcctcc cgtcatgaa 2340
cagtttcaac ctgaaggata tggccccagg gttgggcagt gagacgcggc tggaccgcag 2400
caaaggagat gcacggagcg acacagttct gctcgactcc tcggccacac tcatcacc 2460
tgaggataac gaggaaccgc agggcagcga cgtacctgcc gactacctgc tgggtgacgt 2520
ggaagggtgag gaggacgagc tgtacctgct ggatccggaa cagcgccacc ccgaggacce 2580
tgagtgcgtg ctgccgcagg aggcctttcc gctgcgccac gagatcgtgg acacgcctcc 2640
cgggcccgcg cacctgcagg acaaggccga ctccccgcac gtgagcggca ggcaggcgga 2700
tgttgccctc ctggccccgc tggactctc ctctctgctc ctgtctgtct cacacgcgct 2760
ctacgacagc cgcctcgccg ccgacttctt cggcgtgctg gtgcgcgaca tgctgcact 2820
ctacgctgag cagggcgagc tgcagatggc tgtgtctgtg ctcatcgtcc tgggtgaacg 2880

```

```

gggtgcgcaag gacatcgacg agcagaccca ggagcactgg tacacttcct acatcgacct 2940
gctgcagcgc ttccgcctct ggaacgtgtc caacgaggtg gtcaagctga gcaaccagccg 3000
cgccgtcagc tgcctcaacc aggcctccac caccctgcac gtcaactgca gccactgcaa 3060
gcggcccatg agcagccggg gctgggtctg cgacaggtgc caccgtgcg ccagcatgtg 3120
tgccgtctgc caaccgtag tcaagggtct ctctgtgtgg tgccagggct gcagccacgg 3180
cggccacctg cagcacatca tgaagtggct ggaaggcagc tccactgtc ccgcaggctg 3240
cggccacctc tgcgagtact cctgacgggg catctgctgg gcttgcccgg gcggccgcgt 3300
gcagaacccc gcctggctct gtgcccga ga cgggaggagg ctggaacttg agacctcaat 3360
aaaggaagta gagccgtgt cggcactgcg tgtcacccca tagcgcggcg ggctcgccgt 3420
aggcccgcg ggccagctggc gggcgggggc ttccggcggg gttgtcgcca tggtaaccgg 3480
cggcgcccat agtcgctctc gggcgggcct tccagtgcc tccgtagag gcctcttccg 3540
ggcggggcca aaggagggtca ctccgggggt ccgtccgggt tccggccgt t 3591

```

&lt;210&gt; 38

&lt;211&gt; 3685

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1773215CB1

&lt;400&gt; 38

```

gtcgccgcct ggccggggcg agacgcgggt gcagagcccg cgccggcgctg gaagcgagtg 60
gcggagcggc gggacctcgg cggactcgcc atggaggagg aggggtgtgaa ggaagccggt 120
gagaagcctc ggggagcaca gatggtggac aaggctggct ggatcaagaa gagcagtggg 180
ggcctcctgg gtttctggaa agaccgatat ctgctcctct gccaggccca gctgctgggtc 240
tatgagaatg aggatgatca gaagtgtgtg gagactgtgg agctgggcag ctatgagaag 300
tgccaggacc ttctgtccct cctcaagcga aaacaccgct ttatcctgct gcgatcccca 360
gggaacaagg tcagcgacat caaattccag gcacccaccg gggaggagaa ggaatcctgg 420
atcaaagccc tcaatgaagg gattaaccga ggcaaaaaca aggccttctga tgaggtaaag 480
gtggacaaga gctgcgccct ggagcatgtg acacgggacc ggggtgcgagg gggccagcga 540
cgccggccac caacgagagt ccacctgaag gaggtggcca gtgcagcttc tgacgggtctt 600
ctgcgcctgg atcttgatgt tccggacagt gggccaccag tgtttgcccc cagcaatcat 660
gtcagtgaag cccaacctcg ggagacaccc cggcccctca tgcctcctac caagccttct 720
ctagcacctg agaccaccag ccctggtgac aggggtggaga cccctgtggg ggagagagcc 780
ccaaccctg tctcagcaag ctctgaggtc tcccttgaga gccaaagagga ctacagagacc 840
ccagcagagg aggacagtgg ctctgagcag cctcccaaca gcgtcctgcc tgacaaactg 900
aaggtgagct gggagaaccc cagccccag gagggccctg ctgcagagag tgcagaaccg 960
tcccagggac cctgtctga gacttctgag ctgccccca gggagggtgg gaagccccct 1020
acacccccac ccaagatctt atcagagaaa ctgaaagcct ccatgggtga gatgcagcct 1080
tctgggccac ctgctccagg cacagtgcag gtctcagtga atggcatgga tgacagtcct 1140
gagcctgcc aagcctctca ggctgagggc accccaggaa ctctccaaa ggatgcaaca 1200
acatccacag cactgcccc ctgggacctg ccacctcagt tccatccccg ctgctcctcc 1260
cttggggact tgcttgggga aggcccgcg catcccttgc agcccaggga acggctatat 1320
cgggcccagc tggagggtgaa ggtggcctcg gaacagacgg aaaaactgtt gaacaagggtg 1380
ctgggcagtg agccggcccc tgtagtgcc gaaacattgc tcagccaggc tgtggagcag 1440
ctgaggcagg ccaccaggt cctgcaggaa atgagagatt tgggagagct gagccaggaa 1500
gcacctggc taaggagaa gcggaaggag ctggtgaccc tctacaggag aagtgcaccc 1560
tagggccttc tgggccagag gcacatccc ttctggccat ccatcaagtc catcaaggcc 1620
cagccctgct gagaaatgtg ctctgtcttc tacagcaatg gctgcaggag ggccattggg 1680
catgtcaggg tttggccatg acccgaagag actcctggcg tccctcctac tctgctctgg 1740
ccagtgggtc caggtgccac ccagggtctac tgctggcta tctggcctgg cctctgggct 1800
ggggctgggg ctgggagcac acacgctggg acctatgtgt ttgtgtggtc gttccaaact 1860
gccccagggc tttggggggc gcaacttggg tttctgggaa tgacatcatc tctgttcccc 1920
atccccagta gtttacatct ctgacttctg aatacagcac agctgagccc cctgcagctc 1980
ccatctccag ctattcttag gcaaagagcc tcatggctaa ggcagcctca aagccagccc 2040
ctctcccac ctattctgag tagctgcaga ggccctgggt ccaggctcta ggttcatccc 2100

```

```

tcagttgggg ggaacgtagg acccagctgg agcctcttga gggagatgag aggcctcttt 2160
gtgaggagga cattagctgt gtggcctctc tctctttggc cctgtttcct tttttgcaa 2220
acaaggacat tttctgcagc cccttcctct cagttagcta tgattggagg gcttaggtct 2280
ggaggattca agagtggaag aggaatttaa ggggtccct agtctagtct ctgccctgg 2340
atagtgtcca gccttgata tttctgaaga ggtggatccc agagtggctc tgatgtccac 2400
attagaaaaa cttacttgta atgatcatgt cagccttcag aagagaatcc ccaccaactt 2460
ctgtgcctcc tcagatgggg atttatctgg atctctgtgg ttccttctca gccgaaacag 2520
gtccagtatc ccagtcattt cttcaaatgc tgataggggt atgttggaat ccgaagccac 2580
ttccccgcct tcaagcccca gatgggctgc tctcctgtaa ctttctagga gaagagacat 2640
tttctctctt ccctttcctg gtccatccct gcaccctggc cctctcccag cctctcccc 2700
acattgtccc tgactctagg ggcacatcca gtctccatcg tgctgcagca gctggactga 2760
gggcagagcc tgtaggtgca gaggccctgg ctcccagagt ccagccactc tccctggggc 2820
ctctggggtg agagcagctt ccgataggac ctgccagat ttctgcatgt gcacttttgt 2880
ttactgaaag agagaaaggg gggggtcaca gcaacatgcc ctggccttct tgctctgttc 2940
cccaacccca ctgaggcctg ctgcacaggt caatgccttc gttatcgta ttgtactgtc 3000
actttgttct tgaggtagta gtcaaggatc aggaggggca gatgtcttct ctgggctgcg 3060
tggggccgga gcagaggtga gcagcaatgc actggttcgg gagcccccat cagcctcctt 3120
gtgcaaactg ggcccccatg ccacagtctg gctttccctc catctgccc aggacaagag 3180
caagaaggac atcagttgcc cagtcagtgt atccccctgcc atcttgctt aggaacagcc 3240
ttccccacc agcagccatg gctggctggg gctttagcca agccacctac tgccaggaat 3300
tgagacctca gttccctcct gtgtcaagta gctaactgca gcagctggac tgagggcaga 3360
gtctgtgggt gcagagaccc tgcagttagg tcacaggttg aggccagcc actctccctg 3420
gggcctggtg ggtaggcaag tagctctggg gccacctcaa gtgaccaaat gctattaatt 3480
tccatccttt agcaggctgg gccctaggca ggaagctggc ttctgggaga ggagtgaaga 3540
cgtgcagggc ctgcctagct tgctgtcttg aggaaggtgg cattccgtgc ttgcctcctt 3600
gaggaggtg gcattctgtg tcttctgctt atgaagcgcc tttcttaaag tttggcaata 3660
aatccatttt tatggaaaaa aaaaa 3685

```

&lt;210&gt; 39

&lt;211&gt; 3143

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3036986CB1

&lt;400&gt; 39

```

ttcggctcga ggtaagactg cgacggtacc gggggcgccg ggaaggaccg agaggcgga 60
ggagcagcgg ctcaggcgcc tgcaaaactg tgccctgaac gaggtagacc atgactgtgg 120
tttcagtggc gtcaactcgt gggctgctct tccctgaggtt ttcctaagcc atccccctggc 180
ggaaccgccc ccagtttttg tccagaagtg cttatagaaa agatggctaa tattaacct 240
aaagaaataa ccttaatagt aggtgtggtt actgcctgct attggaacag cctcttttgt 300
ggttttgttt ttgatgatgt ttcagcaata ctggataaca aagacttgca tccatctaca 360
cctttaaaaa ctttatttca aaatgacttc tggggaaccc ctatgtctga ggagagaagc 420
cacaagtctt accgtccctt aacagtattg acatttcgct taaattattt gttaagtga 480
ctaaaaccaa tgtcatatca tctcctgaat atgatttttc atgctgtggg tagtgtgata 540
tttctcaaag tatgcaaact ttttctggac aacaagagta gtgtgattgc ttctttactt 600
tttgagtgca acccaatata cacagaagca gtaacaggag ttgttggaag agcagaactt 660
ttgtcatcta tctttttctc agcagctttt ttgtcatata ccagatcaa aggaccagac 720
aattccataa tatggactcc aattgccttg acagtgttt tagtggctgt tgcaacatta 780
tgtaagaac aaggaataac agttgttaga atttgcctgt tgatgaagt gtttattgac 840
caggggtata ctttgccatt actatgtact actgctggac agtttctccg tggaaagggt 900
agcattccat tttctatgct gcagacacta gtaaaactca ttgtcttgat gttcagtaca 960
ttattacttg ttgtgattag agtccagggt attcaatccc aacttccagt attcaccagg 1020
tttgataacc cagctgctgt aagcccaact cctacaaggc aactaacttt taactacct 1080
cttctgtgta atgcttggtt gttattaaat ccttcagagc tctgctgtga ttggaccatg 1140
ggaacaatac cacttataga gtcattacta gatattcgaa atctggccac atttactttc 1200

```

```

ttttgttttc tggggatggt gggagtattc agtatcagat actctggtga ttctccaag 1260
actgttttaa tggcgctttg tttaatggca ttaccattta ttctgcac gaaccttttt 1320
ttccagttg gatttggtgt tgccgagcga gtattatatg ttcccagcat ggggttctgt 1380
attttggtag cccatggatg gcagaaaata tcaacaaaaa gtgtatttaa aaagctatcc 1440
tggatttggt tgtctatggt gatactcact cattccttaa aaacattcca cagaaattgg 1500
gattgggagt ctgaatatac attgtttatg tcagccttga aggtaaataa aaataatgcc 1560
aaactttgga ataattgtggg tcatgctctg gaaaatgaaa agaactttga gagagctttg 1620
aaatacttct tacaggctac ccatgttcag ccagatgata ttggtgccc tatgaatgta 1680
ggaagaactt ataaaaattt aaatagaacc aaagaagctg aagaatctta catgatggct 1740
aaatcactga tgcctcaaat tattcctggt aaaaaatatg cagccagaat tgcccctaac 1800
cacctaaatg tttatatcaa tctggctaac ctgatccgag caaatgagtc ccgactggaa 1860
gaagcagatc agctgtaccg tcaagcaata agcatgaggc ccgacttcaa gcaggcttac 1920
attagcagag gagaattgct tttaaaaatg aataaacctc ttaaagcaaa ggaagcatat 1980
cttaaagcac tagagctgga cagaaataat gcagatcttt ggtacaactt ggcaattgta 2040
catattgaac ttaaagaacc aaatgaagcc ctaaaaaact ttaatcgtgc tctggaacta 2100
aatccaaagc ataaactagc attattcaac tctgctatag taatgcaaga atcaggtgag 2160
gttaaaactc gacctgaagc tagaaaaacga cttctaagtt atataaatga agagccacta 2220
gatgctaatt ggtatttcaa tttgggaatg cttgccatgg atgacaaaaa ggacaatgaa 2280
gcagagatgt ggatgaagaa agccataaag ttacaagccg acttccgaag tgctttgttt 2340
aatctggctc tcctgtattc ccagactgca aaggaattaa aggctttgcc aattttggag 2400
gagttactca gatactaccc tgatcatatc aagggcctca ttttaaagg agacattctg 2460
atgaatcaaa agaaagatat actaggagca aaaaaatggt ttgaaaggat tttggagatg 2520
gatccaagca atgtgcaagg aaaacacaat ctttgtgttg tttattttga agaaaaagac 2580
ttattaaaag ctgaaagatg ccttcttgaa acactggcat tagcaccaca tgaagaatat 2640
attcagcgcc atttgaatat agtcagggat aagatttcct catctagttt tatagagcca 2700
atattcccaa ccagtaagat ttcaagtgtg gaaggaaaga aaattccaac tgaaagtgta 2760
aaagaaatta gaggtgaatc cagacaaaca caaatagtaa aaacaagtga taataaaagt 2820
cagtctaaat ccaacaaaca attaggaaaa attggagacg aagagacacc ccacaaaaca 2880
acaaaagaca tcaaagaaat tgagaagaaa agagttgctg ctttaaaaag actagaagag 2940
attgaacgta ttttaaatgg tgaataacat taatatttat cgtgacaatg gtatcaaaga 3000
acatcaatcc gtatcatgtg attgctttta ctgggagctt tgaaaaaaag ttcaagggtt 3060
cctaattggtc aatcatgagc tgccttgaag taggatcaaa ataagatttt cattaaagac 3120
ctgtattatc ccaaaaaaaa aaa 3143

```

&lt;210&gt; 40

&lt;211&gt; 1759

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2041080CB1

&lt;400&gt; 40

```

ggggctctgc gcgagcgagg gacgacggaa gggacgggca ggtgtgggag cggggccacg 60
cagcccgcagc gcgggagtcg caggtgctgg gtgcatgggc cagtgaggac gcacagagat 120
ccctcgccgc gcgaggagg agcagcgagg gagccaggcg ctgcccgaag accctgcctg 180
cgtccgagcg agcggaacct cgcgcttcgc ccggggacaa tccgaagtcc gcgctatgga 240
agaggagaaa tatttgctcg agctgatggc agagaaagat agcctggatc catctttgt 300
gcatgcgtcg cgcottttgg cagaagaaat tgaaaagttt caagggttctg atggaaaaaa 360
ggaagacgaa gaaaagaagt atcttgatgt catcagcaac aaaaacataa agctctcaga 420
aagagtactg attcctgtca agcagtatcc aaagttcaat tttgtgggga aattgcttgg 480
accaagagga aactccttga agaggctaca ggaagaaaca ggtgctaaaa tgtctatcct 540
gggcaaagga tcaatgagag ataaagctaa ggaagaagaa ctaaggaga gtggggaagc 600
caaatatgcc cacttgagtg atgagcttca tgtattaatt gaagtgtttg ctccacctgg 660
ggaagcttat tcacgtatga gtcatgcatt ggaagagatt aaaaaattcc tgggttcctga 720
ctacaatgat gaaattcgctc aggaacaact acgtgaatta tcttacttaa atggctcaga 780
ggactctggt cgtggcagag gtattagagg cagaggggac agaatagctc ccacagctcc 840

```

```
ttcaaggggc cgtgggggtg ccattcctcc tccccacca cctggacgag gtgttctcac 900
ccctcgggga agcactgtaa cccgtggagc gcttccagtg ccacctgtag caagaggtgt 960
ccctaccctt cgagcccggtg gggcaccaac agtgccagga tacagggcac ctctcctcc 1020
agcccatgaa gcttatgaag aatatggtta tgatgatggc tacgggggtg aatatgatga 1080
ccagacctat gagacttatg ataacagcta tgcgaccaa acacaaagtg tgcctgaata 1140
ctatgactac ggtcatggag taagtgagga tgcctatgac agctacgcac cagaagaatg 1200
ggccacaacc cgctctagct tgaaggcacc accgcagagg tcagccagag ggggatacag 1260
ggaacacccc tatggtagat attgaaggto cttccacact gtgacctcac ctcaaagaca 1320
attcatagcc tgtggtctcc acataaacag caacaagaca agtaccaca aacaaagaac 1380
acataattga ataaccgatg tgatttgcta agaaggatgg aggtgaagat gtatgataga 1440
aaaacgaaaa actggaataa acccaagact ggctgtttgg aggtttaatg tttttaacaa 1500
tggaggccaa aaaagaagat aatttggatt tgaaccaata ttttcttcat ctcaactcac 1560
ctcaccacta cagatccaag cctgacacag agcctaacag aaatactttc ctccaaatac 1620
aatcctgttc accatggaag acattaaaac cttcagtggg tgctgctggg ggccttcacc 1680
aaaatgcagc cattcacctt gctgctgggg gccttcacca aaatgcagcc attcaccttc 1740
tttgcttttg ctctaccta 1759
```